
MEGI

MESTRADO

Estatística e Gestão de Informação

***OS ERROS NÃO AMOSTRAIS E OS FATORES DA
NÃO-RESPOSTA EM INQUÉRITOS POR
AMOSTRAGEM NA CIDADE DE MAPUTO***

Basílio Sozinho Cubula

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Estatística e Gestão de
Informação

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

OS ERROS NÃO AMOSTRAIS E OS FATORES DA NÃO-RESPOSTA EM INQUÉRITOS POR AMOSTRAGEM NA CIDADE DE MAPUTO

por

Basílio Sozinho Cubula

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização em Análise e Gestão de Informação

Orientador: Professor Doutor Jorge Morais Mendes

Fevereiro 2013

DEDICATÓRIA

Dedico este trabalho aos meus pais (em memória) que muito cedo partiram
antes de colher o que plantaram.

Que Deus lhes dê um descanso eterno!

AGRADECIMENTOS

O meu primeiro agradecimento é a Deus, pela vida, proteção, saúde e, toda a Sua graça que me permitiu enfrentar todos os desafios da caminhada que hoje é concluída com muita satisfação e alegria.

Em seguida, o meu profundo agradecimento vai para o Professor Doutor Jorge Mendes, por ter aceitado a orientação desta dissertação. Agradeço pela liberdade que me proporcionou, pelo acompanhamento e pelas críticas e sugestões construtivas que tive o privilégio de receber.

A minha esposa Ana Aventina, aos meus filhos Erwin de Annat, Hernandes Ivan, Eliseu Inácio, que ficaram privados do meu total carinho, agradeço pelo grande significado que vão dar a este trabalho.

Um agradecimento especial é dirigido ao INE de Moçambique pela oportunidade dada e pela disponibilização de informação para enriquecimento desta dissertação.

Ao estado português, em particular ao Instituto Português de Apoio ao Desenvolvimento, vão os meus sinceros agradecimentos pelo financiamento concedido.

Meus agradecimentos estendem-se aos colegas do ISEGI, em particular para Luís Pedro, Marlon Francisco, Vania Picanço, Ricardo Godinho, Teresa Lopes, Susana Rodrigues, pelo encorajamento e apoio moral.

A todos os familiares, colegas e amigos que direta ou indiretamente contribuíram para a concretização deste trabalho, o meu Muito Obrigado!

RESUMO

A Estatística desempenha um papel de relevo na transformação de informação em conhecimento. E a qualidade dos resultados depende da qualidade da informação recolhida. Em estudos por amostragem podem ocorrer erros amostrais (inerentes a amostragem em si) e os não amostrais – estes são objeto desta dissertação. Os erros não amostrais são alheios à amostragem e podem ocorrer nas diversas etapas do processo do inquérito. Os vários tipos de erros não amostrais podem ser sistematizados em erros de Especificação, de Base de Sondagem, de Não-resposta, de Medida e de Processamento.

Os erros de não-resposta têm sido uma das preocupações dos pesquisadores, pois influenciam as taxas de resposta, principalmente em populações humanas. A literatura defende que quanto maior for o nível de urbanização maior é o número de não-respostas. Na Cidade de Maputo, onde o nível de urbanização é cada vez mais elevado, os resultados desta dissertação indicam que a taxa de resposta aos inquéritos realizados pelo INE em 2003, 2004 e 2009 (respetivamente IDS, IFTRAB e IOF), está abaixo dos 90% e tende a diminuir. Apesar da semelhança da metodologia empregue na recolha de dados destes inquéritos, ainda os resultados do presente estudo mostram que a taxa de resposta depende do tipo de inquérito. Os agregados familiares tendem a cooperar em inquéritos ligados à força de trabalho do que naqueles cujos conteúdos estão ligados à saúde, às despesas ou rendimentos das pessoas. Estudos pormenorizados podem apurar as reais causas deste comportamento dos agregados familiares.

Os fatores da não-resposta identificados no presente estudo e que concorrem para redução da taxa de resposta nos inquéritos na Cidade de Maputo são a recusa total, a ausência, a incapacidade ou inaptidão dos respondentes, a recusa durante a entrevista e outros fatores ligados à infelicidade e barreiras de língua. Destes, os principais são a recusa total e a ausência dos agregados familiares das suas habitações no período de entrevistas.

PALAVRAS-CHAVE

Inquérito por amostragem, Erros não amostrais, Taxa de não-resposta, Enviesamento de não-resposta

ABSTRACT

Statistics plays a key role in the transformation of information into knowledge. And the quality of the results depends on the quality of the harvested information. In sampling surveys, sampling errors may occur (inherent to the sampling process itself) and also nonsampling errors – these are the object of this dissertation. Nonsampling errors are alien to the sampling process, and may occur within the diverse stages of the survey process. The various nonsampling error types may be systemised as Specification errors, Frame errors, Nonresponse errors, Measurement errors and Processing Errors.

Nonresponse errors have been one of the concerns of researchers, since they influence response rates, especially in human populations. Literature defends that, the higher urbanization is, the bigger is nonresponse. In Maputo city, where urbanization is increasingly higher, the findings of this dissertation indicate that response rate to surveys conducted by INE in 2003, 2004 and 2009 (respectively IDS, IFTRAB and IOF), is under 90% and dropping. Despite the methodological similarity in these surveys data collection, still the results of this study show that the response rate depends on the type of survey. Family households tend to cooperate with surveys related with labour force, more than the ones connected to health, expenses or people income. Detailed studies may ascertain the real causes to this households' behaviour.

Identified nonresponse factors in this study that contribute to the reduction of the response rate in Maputo city, are the complete refuse, absence, respondents incapacity or ineptitude, absolute refuse during interview along with other factors related to unhappiness, and language barriers. Within these, the most important are complete refuse and absence of the whole household members during the interviews schedule.

KEYWORDS

Sampling survey, Nonsampling errors, Nonresponse rate, Nonresponse bias

ÍNDICE GERAL

ÍNDICE DE FIGURAS.....	x
ÍNDICE DE GRÁFICOS.....	xi
ÍNDICE DE QUADROS	xii
LISTA DE SIGLAS E ABREVIATURAS	xiv
1. INTRODUÇÃO	1
1.1. Discussão do Contexto e Identificação do Problema.....	2
1.2. Questões de Investigação	5
1.3. Objetivos do Estudo	6
1.4. Importância e Relevância do Estudo.....	7
2. REVISÃO DA LITERATURA.....	9
2.1. Erros Amostrais e Não Amostrais	9
2.1.1. Erros de Cobertura	11
2.1.2. Erros de Medida	12
2.1.3. Erros de Processamento.....	13
2.1.4. Erros de Não-Resposta	13
2.2. Erro Total num Inquérito por Amostragem e suas Componentes.....	16
2.3. Medição do Erro de Resposta	19
2.4. Medição do Erro de Não-Resposta	20
2.5. Estimadores de Erros Não Amostrais em Entrevistas e Reentrevistas	22
2.6. Taxas de Resposta em Inquéritos por Amostragem	26
2.6.1. Cálculo da Taxa de Resposta	30
2.6.2. Fatores que afetam a Taxa de Resposta	33
2.6.3. O Enviesamento devido à Não-Resposta	37
3. METODOLOGIA DO ESTUDO	42
3.1. Aspetos Gerais da Metodologia usada	42
3.2. Procedimentos Metodológicos do IDS2003.....	44
3.2.1. Objetivos do IDS2003	44
3.2.2. A Amostragem do IDS2003	45
3.2.3. Listagem de Agregados familiares do IDS2003	47
3.2.4. Questionários do IDS2003	48
3.2.5. Organização do Trabalho de campo do IDS2003	49

3.2.6. Procedimentos aplicados na Condução de entrevistas do IDS2003	50
3.2.7. Procedimento para Contactos aos agregados familiares.....	53
3.3. Procedimentos Metodológicos do IFTRAB2004	56
3.3.1. Objetivos do IFTRAB2004	56
3.3.2. Dimensão da Amostra do IFTRAB2004	57
3.3.3. Listagem de agregados familiares do IFTRAB2004	58
3.3.4. Questionários do IFTRAB2004.....	58
3.3.5. Organização do trabalho de campo do IFTRAB2004.....	59
3.3.6. Procedimentos usados na Condução de entrevistas do IFTRAB2004.....	59
3.3.7. Procedimento para Contactos aos agregados familiares.....	60
3.4. Procedimentos Metodológicos do IOF2009	61
3.4.1. Objetivos do IOF2009	61
3.4.2. A Amostragem do IOF2009	62
3.4.3. Dimensão da amostra do IOF2009	62
3.4.4. Listagem de agregados familiares do IOF2009	63
3.4.5. Questionários do IOF2009.....	63
3.4.6. Organização do trabalho de campo de IOF2009.....	64
3.4.7. Procedimentos aplicados na Condução de entrevistas do IOF2009.....	65
3.4.8. Procedimento para Contactos aos agregados familiares.....	65
3.5. Técnicas e Métodos de Análise dos Dados	65
3.5.1. Semelhanças nos Planos Amostrais dos três inquéritos	66
3.5.2. Análise das BD's e os Testes de Significância	67
4. ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS	71
4.1. Descrição das Bases de Dados usadas	71
4.1.1. Base de Dados do IDS2003	72
4.1.2. Base de Dados do IFTRAB2004.....	77
4.1.3. Base de Dados do IOF2009.....	82
4.2. Análise Comparativa dos Dados.....	87
4.3. Discussão dos Resultados	90
4.3.1. As Taxas de Resposta por inquérito	90
4.3.2. Os Fatores de Não-Resposta por inquérito	92
5. CONCLUSÕES E RECOMENDAÇÕES.....	94
5.1. Principais Conclusões	94
5.1.1. Fontes de Erros Não Amostrais	94

5.1.2. Taxa de Resposta e Principais Fatores da não-resposta	95
5.2. Recomendações	97
5.2.1. A nível institucional	98
5.2.2. Modo e condições definidas do inquérito	98
5.2.3. A nível do questionário usado	99
5.2.4. A nível dos inquiridores vs respondentes	99
5.3. Limitações e Proposta de Estudos Futuros	100
6. BIBLIOGRAFIA E REFERÊNCIAS	102
ANEXOS	107

ÍNDICE DE FIGURAS

Figura 2. 1 - Quadro de perda de dados por não-resposta no processo do inquérito por amostragem (Barriball e While, 1999:678).	16
Figura 2. 2 – Taxas de Recusa nos Inquéritos sobre Atitudes do Consumidor, 1954-1976. (Steeh, 1981:45).....	27
Figura 2. 3 – Disposição final da amostra com as respectivas unidades componentes de respondentes e não respondentes (Biemer e Lyberg, 2003:85).	29
Figura 2. 4 – Processo de pesquisa através da internet (Fan e Yan, 2010:133).	34
Figura 4. 1 – Uma parte das variáveis da BD de AF's do IDS2003, em SPSS.....	72
Figura 4. 2 – Uma parte das variáveis da BD de AF's do IFTRAB2004, em SPSS.....	77
Figura 4. 3 – Uma parte das variáveis da BD de AF's do IOF2009, em SPSS.	82
Figura 4. 4 – Uma parte das variáveis da BD conjunta dos três inquéritos, em SPSS	87

ÍNDICE DE GRÁFICOS

Gráfico 4. 1 – Distribuição percentual da variável ENTRCOMPLET do IDS2003.....	75
Gráfico 4. 2 – Distribuição percentual da variável ENTRCOMPLET do IFTRAB2004. Fonte: Elaboração própria	80
Gráfico 4. 3 – Distribuição percentual da variável ENTRCOMPLET do IOF2009.....	85
Gráfico 5. 1 – Taxas de resposta por inquérito e sua significância.....	96
Gráfico 5. 2 – Fatores de Não-Resposta por inquérito	97

ÍNDICE DE QUADROS

Quadro 2. 1 - Atribuições aos inquiridores e supervisores no inquérito principal e nas reentrevistas (Lyberg et al., 1997:735).....	24
Quadro 3. 1 - Dimensão e repartição da Amostra do IDS2003. Fonte: INE-M, IDS2003.	46
Quadro 3. 2 - Dimensão e repartição da Amostra do IDS2003. Fonte: INE-M, IDS2003	47
Quadro 3. 3 - Dimensão e repartição da Amostra do IFTRAB2004. Fonte: INE-M, IFTRAB2004.....	57
Quadro 3. 4 – Dimensão e repartição da amostra do IOF2009. Fonte: INE-M, IOF2009	63
Quadro 4. 1 – Dimensão das amostras dos três inquéritos na Cidade de Maputo. Fonte: IDS2003, IFTRAB2004, IOF2009	71
Quadro 4. 2 – Resultados das visitas ao AF para entrevistas do IDS2003. Fonte: Elaboração própria.....	73
Quadro 4. 3 – Frequências de AF's segundo as visitas efetuadas durante o IDS2003. Fonte: Elaboração própria	73
Quadro 4. 4 – Fatores de Não-Resposta no IDS2003	76
Quadro 4. 5 – Distribuição de frequências esperadas dos não respondentes no IDS2003	76
Quadro 4. 6 – Frequências de AF's segundo as visitas efetuadas durante o IFTRAB2004. Fonte: Elaboração própria	78
Quadro 4. 7 – Resultados das visitas ao AF para entrevistas do IFTRAB2004. Fonte: Elaboração própria.....	78
Quadro 4. 8 – Fatores de Não-Resposta no IFTRAB2004	81
Quadro 4. 9 – Distribuição de frequências esperadas dos não respondentes no IFTRAB2004.....	81
Quadro 4. 10 – Frequências de AF's segundo as visitas efetuadas durante o IOF2009. Fonte: Elaboração própria	83
Quadro 4. 11 – Resultados da visita ao AF para entrevistas do IOF2009. Fonte: Elaboração própria.....	83

Quadro 4. 12 – Fatores de Não-Resposta no IOF2009	85
Quadro 4. 13 – Distribuição de frequências esperadas dos não respondentes no IOF2009.....	86
Quadro 4. 14 – Agregados entrevistados segundo o número de visitas.....	87
Quadro 4. 15 – Casos de Não-Resposta por inquérito segundo seus fatores	88
Quadro 4. 16 – Distribuição de frequências para teste de independência do Qui-quadrado.....	89

LISTA DE SIGLAS E ABREVIATURAS

AAPOR	<i>American Association for Public Opinion Research</i>
AE	Área de Enumeração
AF	Agregado Familiar
AL	Agente de Listagem
BD	Base de Dados
FMI	<i>Fraction of Missing Information</i>
IAF	Inquérito aos Agregados Familiares
IDS	Inquérito Demográfico e de Saúde
IFTRAB	Inquérito à Força do Trabalho
INE-M	Instituto Nacional de Estatística de Moçambique
IOF	Inquérito ao Orçamento Familiar
SPSS	<i>Statistical Package for Social Sciences</i>
UPA	Unidade Primária de Amostragem
USA	Unidade Secundária de Amostragem

1. INTRODUÇÃO

A Estatística enquanto ciência teve as suas origens na organização das atividades do Estado, e desde os seus primórdios criou métodos de estruturação de dados, que influenciaram outras ciências (Aubyn e Venes, 2011). Ainda de acordo com estes autores, a investigação científica contemporânea assenta as suas bases nas ideias criadas pela Estatística, recolhendo os dados através de planos de experiências e de métodos de amostragem sofisticados. Gerações de cientistas, utilizando os meios tecnológicos recolhem, processam, analisam dados e disseminam os resultados da sua investigação com objetivo de apoiar a tomada de decisão.

Segundo Aleixo *et al* (2011), a Estatística desempenha um papel de relevo na transformação de informação em conhecimento; e a qualidade dos resultados depende inevitavelmente da qualidade da informação recolhida. O papel da amostragem, nesse contexto, é relevante no controlo dos erros que surgem na condução de operações estatísticas. Mais precisamente, nas diversas estratégias de obtenção de amostras aleatórias e conhecida a dimensão que a amostra deve ter para garantir a precisão e grau de confiança requeridos. Há, no entanto, erros que de um modo geral só podem ser parcialmente controlados, e que nada têm a ver com a variabilidade dos estimadores baseados em amostras aleatórias.

Segundo Cochran (1977), o processo de condução de estudo por amostragem percorre etapas e cada uma delas com tarefas complexas. Neste processo podem ocorrer erros de duas naturezas: amostrais e não amostrais.

O presente trabalho está inserido no âmbito da dissertação de Mestrado em Estatística e Gestão de Informação. O seu objetivo principal é descrever os erros não amostrais em geral e, particularmente, identificar os principais fatores dos erros de não-resposta que afectam a taxa de resposta aos inquéritos por amostragem. Um estudo empírico vai ser conduzido na Cidade de Maputo tendo em vista a análise das

taxas de resposta em três inquéritos nacionais realizados pelo Instituto Nacional de Estatística de Moçambique (INE-M), designadamente, o Inquérito Demográfico e de Saúde de 2003 (IDS2003), o Inquérito à Força do Trabalho de 2004/2005 (IFTRAB2004) e o Inquérito ao Orçamento Familiar de 2009 (IOF2009).

A elevada heterogeneidade dos habitantes da Cidade de Maputo favoreceu a escolha deste espaço territorial, pois os habitantes da Cidade de Maputo apresentam características sócio-demográficas e económicas diversificadas e completamente diferentes das dos habitantes das restantes províncias moçambicanas (INE, Indicadores Básicos 2011)¹. Essa heterogeneidade pode afetar o grau de participação e colaboração em operações estatísticas. Ainda segundo os Indicadores Básicos (2011), a Cidade de Maputo onde vai ser realizado este estudo possui seus habitantes com as melhores condições sócio-económicas, maior acesso a informação veiculada pelos canais de comunicação social sobre os diversos temas do dia-a-dia comparativamente às restantes províncias moçambicanas. Por outro lado, a Cidade de Maputo foi escolhida para este trabalho por ser um dos domínios de estimação de resultados dos inquéritos por amostragem realizados pelo INE-M.

1.1. DISCUSSÃO DO CONTEXTO E IDENTIFICAÇÃO DO PROBLEMA

Os inquéritos por amostragem podem ser conduzidos segundo um plano probabilístico ou não probabilístico. Segundo Sarndal (1992), a abordagem do plano probabilístico assume idealmente que sejam levadas a cabo as seguintes ações na implementação das operações do inquérito:

- (a) Construir uma base de sondagem perfeita da população alvo;
- (b) Definir a mecânica de seleção da amostra desta base de sondagem com vista a salvaguardar o desenho da sondagem probabilística;

¹ No Anexo A1 apresenta-se o Quadro de indicadores demográficos básicos produzidos com base nos resultados dos Censos de 1997 e 2007 de Moçambique.

- (c) Observar o verdadeiro valor de cada unidade e em todas as unidades da amostra selecionada;
- (d) Processar os resultados sem introdução de erros;
- (e) Usar os dados processados e corretos para fazer inferências válidas.

Quando estas condições ideais são salvaguardadas na sondagem probabilística, o erro amostral é somente derivado da diferença existente entre a estimativa amostral e o verdadeiro valor do parâmetro na população. Assim, segundo Sarndal (1992), algumas questões se colocam:

- Como se pode tomar em consideração os erros introduzidos pela imperfeição das operações do inquérito na influência estatística que, até agora reconhecida somente as variações amostrais?
- Como pode a variância ou a componente do enviesamento causada pelos erros não amostrais ser estimada?

Porque atualmente não há uma teoria completa dos erros não amostrais e porque não se tem conhecimento ordinariamente suficiente acerca da distribuição dos erros não amostrais, estas questões não são habitualmente bem respondidas, isto é, estas questões não têm respostas acabadas (Sarndal, 1992). Este autor salienta que para obter resultados precisos é necessário que cada uma das duas componentes do erro seja pequena.

Outra preocupação que tem sido levantada tem a ver com as taxas de resposta. A taxa de resposta em inquérito por amostragem corresponde a proporção entre as entrevistas (respostas) completas e a soma do número de todos os casos em que houve entrevistas completas, entrevistas iniciadas mas não completadas, recusas, não se estabeleceu contacto com a unidade amostral selecionada, outros casos em que a unidade amostral que deveria ter sido inquirida não o pôde ser (por exemplo em populações humanas, a unidade selecionada pode ter falecido, incapaz de responder por razões físicas ou psíquicas, problemas linguísticos, ou ainda outros casos

reportados como erros não amostrais). Algebrica e genericamente, a taxa de resposta é calculada usando-se a fórmula (AAPOR, 2011):

$$Taxa\ de\ Resposta = \frac{Unidades\ com\ Respostas\ Completas}{Total\ de\ Unidades\ Amostrais} \quad (1.1)$$

Conforme a Equação (1.1), a taxa de resposta num inquérito por amostragem depende do total de unidades amostrais que forneceram respostas completas e do total de unidades amostrais que foram definidas no plano de sondagem associado ao estudo.

Num inquérito por amostragem, o desejável é que o número de unidades com respostas completas seja igual ao total de unidades amostrais planificados para o estudo, ou seja, ausência de erros não amostrais de não-resposta. Neste caso ter-se-ia uma taxa de resposta igual a 100 por cento.

Em Moçambique, a tarefa de recolha, processamento, análise e disseminação da informação estatística oficial está ao cargo do INE-M. Desde a sua criação em 1996 o INE-M tem vindo a realizar estudos de âmbito nacional com representatividade provincial – o domínio mais baixo de estimação dos resultados - e, como em qualquer estudo estatístico, ocorrem erros de natureza amostral e não amostral.

O INE-M tem publicado resultados de inquéritos nacionais com estimação para diferentes domínios (Nacional, Nacional Urbano, Nacional Rural, Regional Norte, Regional Centro, Regional Sul, Provincial) incluindo uma secção de estimativas de erros amostrais associados aos resultados. Essas estimativas de erros incluem, em geral, a estimativa do parâmetro, o erro padrão do estimador, o coeficiente de variação do estimador, o intervalo de confiança de 95 por cento para o parâmetro, a estimativa do efeito de sondagem e o número de observações contabilizadas no cálculo da estimativa do parâmetro.

No que diz respeito aos erros não amostrais, o quão afetaram os resultados, normalmente o leitor e utilizador de dados estatísticos não têm esta informação – ou seja, não se avalia, por exemplo, o impacto dos erros não amostrais na taxa de resposta em inquéritos cuja estimativa dos erros amostrais é apresentada. A dimensão da amostra é calculada e ajustada assumindo a taxa de resposta mínima de 90 por cento na população dos agregados familiares da Cidade de Maputo. Por outro lado, os erros não amostrais, especificamente os de não-resposta, influenciam a taxa de resposta final do inquérito. Assim, a partir dos três inquéritos anteriormente anunciados, pretende-se identificar os principais fatores que concorrem para a não-resposta na Cidade de Maputo.

1.2. QUESTÕES DE INVESTIGAÇÃO

Estudos realizados por vários pesquisadores, com destaque para o de Steeh (1981) sobre as taxas de resposta, permitiram avaliar as tendências a longo prazo para o erro de não-resposta, por subgrupos urbanos usando a técnica de regressão em séries temporais. As análises feitas por Steeh demonstraram que havia um crescimento substancial das não-respostas devido, principalmente, a recusa dos entrevistados. Os resultados do estudo de Steeh demonstraram ainda que estas tendências estão relacionadas com o nível de urbanização, pois quanto mais alto for o nível de urbanização, as taxas de recusa tendem a ser mais altas. Portanto, atendendo ao crescente grau de urbanização que caracteriza a Cidade de Maputo nos últimos anos há toda a preocupação de avaliar o grau de participação da população da Cidade de Maputo nos inquéritos realizados pelo INE-M.

Assim, com o desenvolvimento da presente dissertação espera-se obter respostas para as seguintes questões de investigação:

- **Quais são os tipos de erros não amostrais que podem ocorrer em diferentes fases do inquérito por amostragem?**

- **O quão variam as taxas de resposta dada a ocorrência de erros de não-resposta em diferentes etapas dos inquéritos por amostragem na Cidade de Maputo?**
- **Quais são os principais fatores dos erros de não-resposta que afetam a taxa de resposta em inquéritos realizados na Cidade de Maputo?**

As questões levantadas cujas respostas se pretendem alcançar com esta dissertação estão relacionadas com o facto de que, segundo Sarndal (1992), idealmente quando se realiza um inquérito por amostragem:

- (a) Deve existir uma base de sondagem atualizada, completa, correta e sem duplicações;
- (b) Deve recolher-se informação junto de todas as unidades seleccionadas para a amostra;
- (c) Toda a informação recolhida deve ser exata e corretamente registada e codificada.

1.3. OBJETIVOS DO ESTUDO

O presente estudo tem os seguintes objetivos:

Objetivo Geral

- Descrever os erros não amostrais e identificar os principais fatores do erro da não-resposta em inquéritos por amostragem realizados na Cidade de Maputo.

Objetivos específicos

- Descrever os tipos de erros não amostrais que ocorrem em inquéritos por amostragem.

- Calcular as taxas de resposta em três inquéritos por amostragem: IDS2003, IFTRAB2004, IOF2009 para a Cidade de Maputo.
- Comparar as taxas de resposta nos três inquéritos supra indicados.
- Identificar os principais fatores que influenciam a não-resposta nestes três inquéritos na Cidade de Maputo.

1.4. IMPORTÂNCIA E RELEVÂNCIA DO ESTUDO

A importância deste estudo prende-se pelo facto de as taxas de respostas aos inquéritos, para além de outras medidas de qualidade, conferirem fidedignidade aos resultados, pois quanto mais baixas forem as taxas de respostas menos fiáveis são os resultados, segundo Dey (1997); Dillman, Eltinge, Groves e Little (2002); Jones (1996) citados no artigo de Porter e Whitcomb (2003). Portanto, torna-se relevante identificar e descrever as diferentes fontes de erros não amostrais e, ainda identificar os principais fatores que influenciam, particularmente, os erros de não-resposta. Esses fatores têm um impacto direto nas taxas de respostas em inquéritos cujos resultados se pretende que sejam estatisticamente válidos e tomados como ferramenta de apoio na tomada de decisão.

Os resultados de estudos estatísticos em geral são disseminados e tidos como válidos para diferentes fins. Contudo, questões ligadas às taxas de resposta, os fatores das não-respostas e outros constrangimentos experimentados na condução do estudo e que medidas foram tomadas para prevenir e controlar a ocorrência da não-resposta, não são tidas em conta pelos produtores de informação estatística e, muito menos, pelos potenciais utilizadores dos resultados. Isto leva-nos a concluir a relevância deste estudo pelo seu carácter que pode ser considerado singular e, de certo modo, importante para as instituições e organizações de pesquisa e estudos de mercado em Moçambique.

A taxa de resposta é um dos indicadores de qualidade (Biemer e Lyberg, 2003). No processo de inquérito por amostragem, a qualidade é requerida. Para Biemer e Lyberg, qualidade pode ser definida simplesmente como “*fitness for use*”. No contexto de inquérito, a qualidade pode ser traduzida como requerimento para que dados estatísticos sejam precisos o mais necessário possível de modo a atingir os fins requeridos. É também nisto de que se reveste a relevância deste estudo.

Os resultados que vão ser apresentados no âmbito desta dissertação contribuirão para que o INE-M e outras instituições possam tomar decisões que conduzam na melhoria dos procedimentos de planificação, recolha de dados e produção de estatísticas com algum nível de qualidade acrescentada. Adicionalmente, ter-se-á mais uma referência sobre a necessidade de se prestar mais atenção aos erros não amostrais, pois medidas podem ser traçadas com vista ao maior controlo e prevenção da sua ocorrência, particularmente os ligados à não-resposta, em todas as etapas do inquérito.

2. REVISÃO DA LITERATURA

Neste capítulo vai ser apresentada uma descrição e caracterização dos erros não amostrais e apresentadas as diferentes abordagens dos autores consideradas relevantes na contextualização do tema em estudo.

Os erros não amostrais ocorrem durante todo o processo de produção estatística e concorrem para a qualidade estatística (Biemer e Lyberg, 2003). Ainda de acordo com estes autores, uma organização ligada aos inquéritos não difere de outra organização no que respeita à necessidade de melhoria contínua. Há necessidade de boa qualidade dos resultados, mas também é necessário que a organização seja ágil e ajuste seus processos de acordo com a nova demanda de utilizadores de informação. Nesse contexto, a qualidade é exigida. A definição de qualidade, a mais geral e amplamente citada é a que foi apresentada por Juran e Gryna (1980) citados por Biemer e Lyberg (2003) que simplesmente definem como *“fitness for use”*. Portanto, há que identificar e descrever as fontes de erros não amostrais com vista a garantir a máxima qualidade possível dos resultados estatísticos.

2.1. ERROS AMOSTRAIS E NÃO AMOSTRAIS

A confiança dos resultados de um inquérito por amostragem é função do erro total que é composto por duas componentes: erros amostrais e erros não amostrais.

Erros Amostrais. Os Erros Amostrais resultam da não observação da totalidade da população. Este tipo de erro está sempre presente num inquérito por amostragem. O papel do estatístico ou investigador consiste em controlar e, se possível conhecer este erro. Numa amostragem probabilística este erro ocorre ao acaso e é mensurável. O erro amostral traduz a diferença existente entre a estimativa amostral e o verdadeiro valor do parâmetro na população (Cochran, 1977).

De acordo com Freitas (2002), os erros amostrais estão associados à amostragem propriamente dita, pois uma sondagem é um processo de recolha de informação junto de uma amostra e de generalização desta informação amostral para a totalidade da população. O erro amostral decorre da própria noção de amostra, como subconjunto de elementos da população alvo. Ainda esta autora reconhece que esta definição determina que existem elementos da população não representados na amostra, e amostras diferentes sobre a mesma população alvo produzem resultados diferentes.

Por existir variabilidade nas estimativas produzidas a partir de amostras diferentes e por a amostra não constituir uma representação perfeita da população, percebe-se que dificilmente os resultados amostrais serão equivalentes aos verdadeiros parâmetros populacionais.

Para Aleixo *et al* (2011), em geral, a expressão “erro amostral” designa o erro decorrente da variabilidade das estimativas de amostra para amostra; enquanto que a expressão “erros não amostrais” a todos os outros que ocorrem na recolha de dados.

Erros Não Amostrais. Estes resultam dos erros que são introduzidos na implementação do inquérito em todas as suas etapas, por exemplo: impossibilidade de observar certas unidades da população, valores mal registados, erros de codificação. Aliás, esta classificação foi apresentada por Sarndal em 1992 que sistematizou-os em quatro subtipos: Erro de cobertura, Erro de medida ou de observação, Erro de não-resposta, Erro de processamento.

Segundo Assael e Keon (1982), os erros não amostrais mais complicados e de difícil controlo são os de não-resposta e os de resposta. Os primeiros, os erros de **não-resposta**, ocorrem quando algumas unidades amostrais não respondem ao inquérito fazendo com que as outras respostas sejam uma representação imperfeita da amostra selecionada. Os **erros de resposta** ocorrem quando as unidades amostrais fornecem respostas imprecisas. Os erros de resposta podem ocorrer por diferentes razões: os

respondentes propositadamente omitem suas verdadeiras respostas, não se lembram perfeitamente da resposta verdadeira, os respondentes estão fatigados, os respondentes são influenciados pelos inquiridores, os respondentes são afetados por uma série de fatores do ambiente onde se encontram.

Assael e Keon (1982) sustentam que o dilema que frequentemente os pesquisadores enfrentam é saber se selecionam uma maior dimensão da amostra de modo a minimizar os erros não amostrais ou concentrar recursos financeiros e esforço numa amostra de dimensão reduzida e, assim, assegurar um melhor controlo aos inquiridores, elevadas taxas de resposta e maior precisão das respostas. Idealmente os pesquisadores concentram esforço em reduzir ambos os erros amostrais e não amostrais simultaneamente. Dados os constrangimentos de custos e tempo, o ideal raramente é realizado. Este dilema é mais complicado pelo facto de que depois de se conduzir o inquérito, os pesquisadores raramente estão dispostos em determinar (medir) o erro total da pesquisa ou comparar as magnitudes das suas componentes. Pois, o erro total da pesquisa e suas componentes podem ser medidos se forem obtidos dados válidos (respostas efetivamente corretas) para todas as unidades amostrais. Estes dados raramente são alcançáveis.

2.1.1. Erros de Cobertura

Os **erros de cobertura** resultam da utilização de uma base de sondagem que não cobre de forma completa e correta a população alvo. Na realidade, é difícil encontrar adequação perfeita entre a população alvo e a base de sondagem. Para Aleixo *et al.* (2011), os erros de cobertura mais comuns são por defeito: parte das unidades da população alvo estão ausentes da base de sondagem que é usada.

Segundo Barriball e While (1999), normalmente não é possível a obtenção de dados a partir de toda a população, o que significa que os investigadores precisam selecionar e recrutar amostras. No entanto, incluindo alguns indivíduos em uma base de sondagem e excluindo outros corre-se o risco de introduzir enviesamento. Algumas

técnicas para assegurar a seleção de amostras representativas foram desenvolvidas. Barriball e While (1999) salientam por exemplo, a estratificação, a aleatoriedade, a segmentação por *cluster* desenvolvidas por Moser e Kalton (1971). Assim, segundo Barriball e While (1999) é responsabilidade de todos os pesquisadores considerarem cuidadosamente a abordagem ou combinação de abordagens para seleção da amostra que irá minimizar o erro de cobertura de acordo com as circunstâncias e o objeto de pesquisa.

2.1.2. Erros de Medida

Erros na medição da unidade de inquirição. Esta designação foi feita por Cochran (1977). O instrumento de medição ou conteúdo do questionário se for o caso, pode estar enviesado ou impreciso. Em populações humanas os respondentes podem não dispor de informação precisa do assunto que está sendo investigado ou, no pior caso, podem fornecer informação incorreta. Igualmente, esta designação aplica-se a erros motivados por certos comportamentos do entrevistador ou do entrevistado ou ainda, por deficiências na concepção do questionário.

Num artigo publicado em 1999 por Barriball e While estes reforçam que textos de diversas pesquisas têm estabelecido uma série detalhadas de estratégias para minimizar o nível de não-contacto e recusas em pesquisas que utilizam tanto o questionário, entrevistadores bem como no contacto *face-to-face* com potenciais respondentes [Champion e Sears (1969), Dillman e Frey (1974), Gordon (1975), Bradburn e Sudman (1979), Miller (1991) todos citados por Barriball e While (1999)]. Tais textos são, em grande parte, baseados no pressuposto de que há muito que os entrevistadores podem fazer para reduzir o erro de medição das unidades amostrais e elevar as taxas de cooperação.

2.1.3. Erros de Processamento

Os **Erros de processamento** ocorrem por falhas humanas ou de *software* durante a edição, a codificação, a digitação ou a tabulação de dados. Os erros de processamento podem ocorrer em diversas fases do estudo (Cochran, 1977).

A fase de tratamento de dados não está isenta de erros. Assim, em particular podem ser fontes de erro as operações de codificação, transcrição (ex. passagem para suporte magnético das respostas disponíveis em papel), imputação, edição. O uso da nova tecnologia CAPI, CATI, CAWI², de programas de edição e regras de validação contribuíram para a redução deste tipo de erros e para o aumento da qualidade dos dados recolhidos (Aleixo *et al.*, 2011).

2.1.4. Erros de Não-Resposta

Os **erros de não-resposta**, segundo Cochran (1977), estão relacionados com a falta de informação de algumas unidades amostrais. Isto pode ocorrer pela falta de supervisão ou, em populações humanas, pela falha (ou falta) na localização de alguns indivíduos ou pela recusa em prestar a entrevista nos casos em que são localizados.

Ainda de acordo com Cochran (1977), no estudo de não-resposta é conveniente imaginar como se a população estivesse dividida em dois estratos, o primeiro constituído de todas as unidades cujas observações poderiam ser obtidas caso a unidade estivesse caído na amostra, o segundo estrato constituído de unidades cujas observações vão ser obtidas. A composição destes dois estratos depende grandemente dos métodos usados para obter as unidades amostrais e delas obter a informação.

Em 1999 os investigadores Barriball e While apresentaram, no seu artigo, uma discussão metodológica e um modelo explicativo sobre os erros de não-resposta.

²CAPI-Computer Assisted Personal Interviewing, CATI-Computer Assisted Telephone Interviewing, CAWI-Computer Assisted Web Interviewing.

Nesse artigo, Barriball e While examinaram algumas implicações da não-resposta e discutiram como esta pode ser contida e gerida e ainda, realçaram a sua complexidade. Para Miller (1991) citado por Barriball e While (1999) sobre a significância da não-resposta afirma que enquanto a pesquisa oferece um método de recolha de dados que pode produzir uma maior validade através do uso de amostras mais representativas, o potencial para não-resposta é uma grande desvantagem. E ainda, de acordo com Hartman *et al.* (1985) citado por Barriball e While (1999), se as taxas de resposta são baixas e ou a não-resposta é sistemática e de alguma maneira correlacionada com as variáveis em estudo, amostra a partir da qual os dados são recolhidos torna-se não representativa. Em tais casos a validade externa fica ameaçada e conclusões válidas não podem ser obtidas a partir desses dados, como igualmente defendem Williamson (1981) e Denzin (1989) citados por Barriball e While (1999).

O interesse metodológico de não-resposta, portanto, surge a partir do seu potencial para introduzir enviesamento no processo da pesquisa, conforme Moser e Kalton (1971) advertiram, citados por Barriball e While (1999), nunca se deve assumir que as diferenças entre os respondentes e não respondentes não existem.

Barriball e While (1999), definem a não-resposta como sendo o fracasso do pesquisador em obter a cooperação de todos os potenciais entrevistados incluídos na amostra líquida³. No entanto, eles reconhecem existir alguma ambiguidade nesta definição por esconder a complexidade do erro da não-resposta, pois há potenciais fontes de não-resposta que é importante que sejam consideradas pelos pesquisadores [Kviz (1977) citado por Barriball e While (1999)]. Steeh (1981), citado por Barriball e While (1999), por exemplo, identifica duas grandes componentes da não-resposta: (1) as recusas dos entrevistados em cooperar no processo de pesquisa e, (2) o não-contacto que faz com que os dados não sejam recolhidos em todas as unidades amostrais.

³ Segundo Barriball e While (1999), a amostra líquida obtém-se excluindo os indivíduos que não satisfazem os critérios de amostragem e contendo apenas os que se pretende obter informação.

Por outro lado, Barriball e While (1999) citam o artigo produzido em 1988 por Aiken no qual examina a não-resposta e também abordou o problema das recusas, mas incluiu em sua definição o problema da não-cobertura e do potencial enviesamento que a exclusão de indivíduos a partir da base de sondagem podem introduzir.

Numa revisão abrangente da não-resposta, Elliot (1991:3) citado por Barriball e While (1999) distinguiu três tipos de não-resposta:

1. Não-cobertura: quando a base de sondagem omite accidental ou deliberadamente algumas unidades populacionais. Para este autor, a não-cobertura é considerada não-resposta por incluir (ou omitir) accidentalmente algumas unidades da população que serão (não serão) selecionados para amostra. Deste modo, a não-cobertura pode induzir um potencial enviesamento dos resultados.
2. Não-resposta da unidade: quando nenhuma informação é recolhida a partir de uma unidade de amostragem devido a, por exemplo, recusa ou não-contacto;
3. Não-resposta sobre algum item: quando a unidade amostrada concorda em participar do estudo, mas informações sobre algum conteúdo da investigação não é recolhida porque, por exemplo, a unidade de amostragem se recusa ou é incapaz de responder a uma pergunta específica ou o pesquisador não faz a pergunta por engano.

Usando a tripla tipologia de Elliot (1991), segundo Barriball e While (1999), como base para uma definição-padrão da não-resposta, pode ser visto que a não-resposta pode ocorrer em diferentes fases do processo de investigação conforme se ilustra na Figura 2.1.

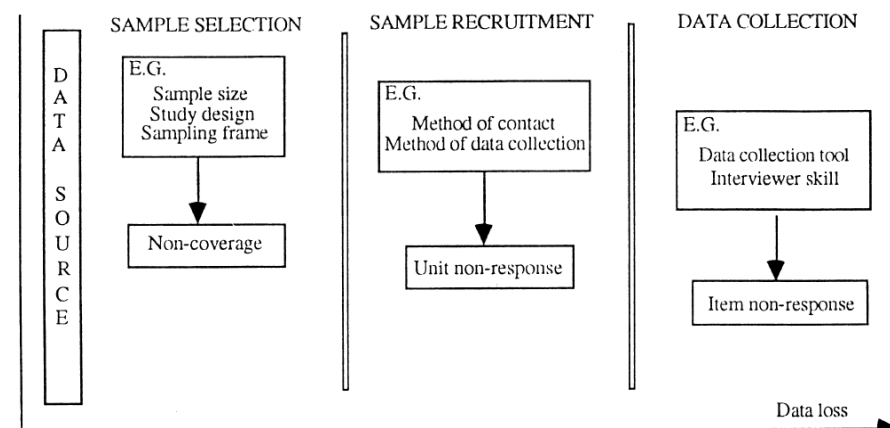


Figura 2. 1 - Quadro de perda de dados por não-resposta no processo do inquérito por amostragem (Barriball e While, 1999:678).

Estas três fontes de falha/falta de resposta, no entanto, podem ser manipuladas por investigadores com vista a minimizar o potencial do erro de não-resposta.

A não-resposta foi também discutida no artigo apresentado por Kreuter, Muller e Trappmann em 2010. Nesse artigo estes autores reconhecem que a não-resposta é um problema incontornável em amostragem. Há sempre quem não esteja acessível, não queira responder a algumas das perguntas ou à totalidade de um questionário, e se a resposta ao questionário é feita de forma não presencial, pedindo, por exemplo, que se devolva pelo correio (usando um sobrescrito fornecido, com selo pago), uma larga fração de inquiridos não responde.

2.2. ERRO TOTAL NUM INQUÉRITO POR AMOSTRAGEM E SUAS COMPONENTES

Considere-se o caso em que o interesse consiste na média de uma característica. De acordo com Raj (1968, Capítulo 8) citado por Assael e Keon (1982), o erro total num inquérito por amostragem é função da diferença entre a média do verdadeiro valor da característica na população onde é extraída a amostra e a média do valor observado na amostra dos respondentes. Deste modo, o erro total pode ser medido como erro

quadrático médio da média amostral dos respondentes à volta do verdadeiro valor médio na população. Ou seja [Raj (1968) citado por Assael e Keon (1982:115)]:

$$\text{Erro Total do inquérito} = f(\bar{X}_{\text{população}} - \bar{y}_{\text{respondentes na amostra}}) \quad (2.1)$$

onde:

$\bar{X}_{\text{pop.}}$ = verdadeiro valor médio do atributo na população

$\bar{y}_{\text{resp.}}$ = valor médio do atributo dos respondentes na amostra

$$\text{EQM Erro Total do inquérito} = \frac{1}{K} \sum (\bar{X}_{\text{pop.}} - \bar{y}_k)^2 \quad (2.2)$$

onde:

$\bar{X}_{\text{pop.}}$ = verdadeiro valor médio do atributo na população alvo

\bar{y}_k = valor médio do atributo observado da amostra k (inclui somente respostas dos respondentes na amostra)

K = número de amostras possíveis que podem ser selecionadas da pop. alvo

Ainda tendo como interesse a média da característica, o erro total do inquérito pode ser dividido em duas componentes – o erro aleatório de amostragem e o erro não amostral. O erro aleatório de amostragem representa quão o verdadeiro valor médio do atributo na amostra selecionada, \bar{x}_{amostra} , é representativo do correspondente verdadeiro valor médio do atributo na população de onde foi extraída a amostra, $\bar{X}_{\text{população}}$. Por outro lado, o erro não amostral indica quão o valor médio do atributo observado para os respondentes na amostra, $\bar{y}_{\text{respondentes}}$, é representativo do verdadeiro valor médio do atributo na amostra, \bar{x}_{amostra} . Assim,

$$\begin{aligned} \text{Erro Total do inquérito} &= \text{Erro Aleatório Amostral} + \text{Erro Não Amostral} \\ f(\bar{X}_{\text{pop.}} - \bar{y}_{\text{res.}}) &= f(\bar{X}_{\text{pop.}} - \bar{x}_{\text{amostra}}) + f(\bar{x}_{\text{amostra}} - \bar{y}_{\text{res.}}) \end{aligned} \quad (2.3)$$

Para o caso dos erros de não-resposta e de resposta, o erro não amostral pode ser definido como função destas duas componentes (Assael e Keon, 1982). O **erro de não-resposta** ocorre porque algumas unidades amostrais podem não responder. Tomando a média da característica como a estatística de interesse, o erro de não-resposta é função de quão o verdadeiro valor médio da característica dos respondentes na amostra, $\bar{x}_{resp.}$, representa o verdadeiro valor médio da característica de todas as unidades amostrais, $\bar{x}_{amostra}$. Por outro lado, o **erro de resposta** ocorre pelo facto de dentre as unidades amostrais que respondem ao inquérito algumas podem fornecer respostas imprecisas ou inadequadas. Portanto, o erro de resposta é função de quão preciso o valor médio da característica observada na amostra, $\bar{y}_{resp.}$, é em relação ao verdadeiro valor médio da característica dos respondentes na amostra, $\bar{x}_{resp.}$.

Erro Não Amostral

$$\begin{aligned}
 &= f(\bar{x}_{amostra} - \bar{y}_{resp.}) \\
 &= f(\bar{x}_{amostra} - \bar{x}_{resp.}, \bar{x}_{resp.} - \bar{y}_{resp.}) \\
 &= f(\text{Erro de Não - resposta}, \text{Erro de resposta})
 \end{aligned} \tag{2.4}$$

Ainda de acordo com Assael e Keon (1982) as estimativas não enviesadas para o EQM do erro total do inquérito (erro aleatório amostral e erro não amostral) todas podem ser obtidas desde que as respostas validadas de todas as unidades amostrais estejam disponíveis ou sejam acessíveis. Infelizmente, não podem ser calculadas simples estimativas do EQM do erro de resposta e do erro de não-resposta que perfaz as estimativas não enviesadas do EQM dos erros não amostrais. Entretanto, podem ser usadas medidas substitutas de erros de resposta e de não-resposta desde que tais medidas sejam consideradas úteis no diagnóstico de fontes de erros não amostrais observados.

2.3. MEDIÇÃO DO ERRO DE RESPOSTA

De acordo com os autores Ferber (1953); Lansing, Ginsburg e Braaten (1961); Metz (1956); Neter e Waksberg (1965) citados por Assael e Keon (1982), o erro de resposta lida com as diferenças entre as respostas dadas pelos respondentes e os valores concretos das respostas aos itens avaliados na pesquisa. Os valores concretos das respostas aos itens avaliados na pesquisa podem ser obtidos somente a partir de validações externas, por exemplo, registros de compras, documentos de identificação, pagamentos de faturas, mapas de registros de acontecimentos ou boletins informativos. Nos inquéritos de qualidade dos dados, pesquisas sobre erros de resposta têm sido reportadas embora os estudos sejam limitados, devido em parte à dificuldades de obter informação externa para validação. Na ausência de validação de cada resposta, alguns métodos têm sido propostos. Os autores Andersen, Kasper e Frankel (1977); Dutka e Frankel (1976); Frankel e Frankel (1977) citados por Assael e Keon (1982), referem que um dos métodos consiste em validar as respostas de uma parte da amostra dos respondentes e, em seguida fazer projeções para o total da amostra ou ainda, o uso de medidas substitutas com propósito de assessorar as avaliações [Brown (1967) citado por Assael e Keon (1982)].

Ainda, os autores Ford (1967, 1968), Houston e Jefferson (1975), Nevin e Ford (1976) citados por Assael e Keon (1982) salientam que quando a validação das respostas a partir de informação externa não é possível, os pesquisadores usam duas abordagens para medir os erros de resposta: (1) o uso de **variáveis proxy**⁴ como item de não-resposta para medida de qualidade de resposta e (2) a determinação de

⁴Variável “Proxy” trata-se de medição indirecta da variável que o pesquisador pretende estudar. É usada quando o objeto de estudo é difícil de medir ou de observar. Uma *variável proxy* substitui aproximadamente a que está-se a estudar. Fonte: <http://pt.scribd.com/doc/15886719/Dicionario-de-tipos-de-variaveis>, acedido em 13 de Maio de 2012.

diferenças nas respostas entre os grupos experimentais emparelhados como medida de enviesamento de resposta.

Assim, no mesmo contexto da média como estatística de interesse, o enviesamento B_r de resposta é definido como sendo a diferença entre o verdadeiro valor médio da característica dos respondentes da amostra para o item pesquisado, \bar{x}_i (as respostas validadas dos não respondentes são excluídas no cálculo desta média) e a média de respostas observadas \bar{y} (Assael e Keon, 1982):

$$B_r = \bar{x}_i - \bar{y} \quad (2.5)$$

Vários estudos têm-se dedicado a redução do enviesamento de respostas e, conseqüentemente no correspondente erro de resposta [Andersen, Kasper e Frankel (1977); Cannell, Oksenberg e Converse (1977); Dutka e Frankel (1976); Frankel e Frankel (1977) citados por Assael e Keon (1982)].

2.4. MEDIÇÃO DO ERRO DE NÃO-RESPOSTA

O erro de não-resposta é outra componente dos erros não amostrais. O erro de não-resposta ocorre quando a existência de não respondentes faz com que os atuais respondentes sejam maus representantes do total da amostra original. Muitos estudos têm sugerido formas de reduzir os erros de não-resposta [Armstrong e Overton (1977); Fillion (1975, 1976); Houston e Nevin (1977); Kanuk e Berenson (1975); Pace (1939) citados por Assael e Keon (1982)]. Frequentemente esses estudos têm usado taxas de resposta como medidas substitutas para os erros de não-resposta. Ainda que a taxa de resposta somente meça o número relativo de não respondentes e ignore as diferenças entre respondentes e o total da amostra. Por isso que a taxa de resposta constitui uma medida substituta pobre do erro de não-resposta. Ou seja, a tarefa de melhorar a taxa de resposta não reduz necessariamente o erro de não-resposta, pois de acordo com

Leslie (1972) citado por Assael e Keon (1982), os não respondentes podem tornar-se cada vez mais diferentes dos respondentes fazendo com que os respondentes sejam piores representantes do total da amostra original.

A melhor medida substituta de erro de não-resposta é o enviesamento da não-resposta validada. No caso em que a estatística de interesse é a média da característica, o enviesamento da não-resposta é definido como sendo a diferença entre o verdadeiro valor médio de todas as unidades amostrais \bar{x} , e o verdadeiro valor médio somente da amostra de respondentes \bar{x}_1 . Esta medida é mais representativa do erro de não-resposta por incluir considerações à volta de ambos os números de unidades que responderam e as diferenças entre as populações de respondentes e não respondentes.

Estas duas medidas substitutas das componentes de erros de não-resposta e todas juntas, resulta no enviesamento não amostral. E, a maior componente estatística de erro não amostral é o quadrado do enviesamento não amostral.

$$\left\{ \begin{matrix} \text{Enviesamento} \\ \text{Não Amostral} \end{matrix} \right\}^2 = \{\bar{x} - \bar{y}\}^2 = \{(\bar{x} - \bar{x}_1) + (\bar{x}_1 - \bar{y})\}^2 = \left\{ \begin{matrix} \left(\begin{matrix} \text{Enviesamento} \\ \text{de Não - res} \\ \text{posta validada} \end{matrix} \right) + \left(\begin{matrix} \text{Enviesamento} \\ \text{de Resposta} \end{matrix} \right) \end{matrix} \right\}^2 \quad (2.6)$$

onde:

\bar{x} = verdadeiro valor médio do atributo em todas as unidades amostrais

\bar{y} = valor médio observado do atributo dos respondentes na amostra

\bar{x}_1 = verdadeiro valor médio do atributo somente dos respondentes na amostra

A título de exemplo, as várias componentes do erro do inquérito foram examinadas num estudo empírico realizado por Assael e Keon (1982) e comparadas para cada uma das demais alternativas de desenho de inquérito por amostragem.

2.5. ESTIMADORES DE ERROS NÃO AMOSTRAIS EM ENTREVISTAS E REENTREVISTAS

Conforme foi exposto anteriormente e de acordo com Bassi e Fabbris (1997), nos inquéritos por amostragem onde os dados são recolhidos por inquiridores a maior fonte de erros não amostrais são os respondentes, inquiridores, supervisores e outros membros da equipa de campo, formadores, codificadores, digitadores. Os respondentes, de facto podem fornecer resposta imprecisa por diversos motivos. Se o desvio da resposta do seu verdadeiro valor tem carácter aleatório resulta num erro chamado **erro de resposta não correlacionado**. Entretanto, se os erros tendem a ser sistemáticos, então um enviesamento de respostas estará presente nos resultados.

O erro dos inquiridores pode ser relacionado com as características pessoais dos inquiridores, as suas atitudes e comportamento durante as entrevistas. O erro resulta num enviesamento dos resultados se for respeitante a um inquiridor particular. Mas, por outro lado, é considerado erro aleatório correlacionado se for em relação a toda amostra extraída da população, i.é, os inquiridores são induzidos a introduzir erros não amostrais. Os erros dos codificadores, digitadores e outro pessoal envolvido na manipulação de dados são estatisticamente similares aos erros introduzidos pelos inquiridores (Bassi e Fabbris, 1997).

Os erros de supervisores é outra fonte de enviesamento de resultados de um inquérito por amostragem (Bassi e Fabbris, 1997). Através da formação ou treinamento do pessoal ou da comunicação, os supervisores influenciam diretamente os inquiridores, e indiretamente influenciam nas respostas recolhidas pelos inquiridores.

Bassi e Fabbris (1997) afirmam que a tal cadeia de formação contínua que tem sido implementada sob instruções dos pesquisadores, que inadvertidamente faz passar informação, atitudes, concepções e comportamentos de uma etapa para outra, mais tarde se transformam em potencial fonte de erro em todo o processo do

inquérito por amostragem. Estes autores chamaram a este tipo de erro não amostral por **erro de instrução**.

Modelo de estudo dos Erros Não amostrais

Em 1997, Bassi e Fabbris propuseram um **modelo de estudo dos erros não amostrais** num inquérito por amostragem assistido por inquiridores e supervisores. Este modelo incorpora, ao mesmo tempo, ambos os efeitos tanto dos inquiridores quanto dos supervisores. No modelo proposto por Bassi e Fabbris (1997) os estimadores das componentes da variância da média amostral e seu erro padrão são propostos e suas eficiências são comparadas. Adicionalmente, o erro quadrático médio (EQM) dos estimadores da variância das respostas correlacionadas com base em entrevistas individuais são comparadas com estimativas que combinam dados da entrevista com os da reentrevista.

Para o modelo proposto por Bassi e Fabbris (1997) considere-se a situação em que um inquérito por amostragem seja conduzido a partir de entrevistas assistidas por inquiridores. Mais tarde são feitas reentrevistas usando uma sub amostra de respondentes para propósitos de avaliação. Suponha-se que a população de dimensão N é dividida em H estratos de onde são extraídas sub amostras sem reposição. Assumindo que cada estrato tem igual dimensão (N_h) e a dimensão da amostra de cada estrato é $n=km$ onde k denota o número de inquiridores e m representa a dimensão atribuída a cada inquiridor. A atribuição dos inquiridores aos estratos é aleatória. Esta atribuição aleatória das unidades aos inquiridores é conhecida por **interpenetração** [Mahalanobis (1994) citado aqui por Bassi e Fabbris (1997)] e é requerida na estimação da variância correlacionada do erro num desenho amostral singular. Assume-se que a mesma interpenetração do inquérito principal é usada na amostra de reentrevistas.

Na reentrevista as atribuições dos inquiridores são trocadas com outros inquiridores dentro do mesmo estrato de modo que cada inquiridor faça reentrevistas

num estrato diferente do seu inicial. Notar que os inquiridores mantêm os mesmos supervisores tanto no inquérito principal como nas reentrevistas. A título ilustrativo, estes autores consideraram o caso de dois supervisores ($H=2$) e quatro inquiridores ($k=2$ em cada estrato). No Quadro 2.1 sob inquérito principal, as letras A, B, C, D representam as atribuições cuja dimensão é m .

	Inquérito principal				Reentrevista			
Supervisor	1		2		1		2	
Inquiridor	1	2	3	4	1	2	3	4
Sub Amostra	A	B	C	D	B	A	D	C

Quadro 2. 1 - Atribuições aos inquiridores e supervisores no inquérito principal e nas reentrevistas (Lyberg et al., 1997:735).

Os pressupostos mantêm-se os mesmos de que existe mais uma componente chamada erro de supervisão. Pode ser demonstrado que este desenho de entrevista-reentrevista é idêntico ao apresentado por Fellegi (1964) citado por Bassi e Fabbris (1997) quando $H=1$.

Com base neste desenho o modelo de resposta pode ser apresentado da seguinte forma para determinada unidade amostral no estrato h , de acordo com Bassi e Fabbris (1997):

$$y_{hij} = \mu_{hij} + b_h + a_{hi} + r_{hij} \quad (2.7)$$

onde

$y_{hij} = (h = 1, \dots, H; i = 1, \dots, k; j = 1, \dots, m)$ é o valor observado da variável aleatória Y normalmente distribuída

μ_{hij} é o correspondente valor verdadeiro ou valor "livre de erro" da unidade j dentro da atribuição i do inquiridor referente ao estrato h

b_h é o erro devido ao h -ésimo supervisor

a_{hi} é o erro devido ao hi -ésimo inquiridor

r_{hij} é o desvio de y_{hij} em relação a $\mu_{hij} + b_h + a_{hi}$, tido como erro de resposta

Assim, assume-se que a resposta é o resultado do verdadeiro valor individual μ_{hij} distorcido pelos erros do respondente, inquiridor, ou supervisor.

Da aplicação deste modelo de avaliação de erros de respondentes, inquiridores, supervisores para determinar estimativas das componentes de erros de resposta do valor da variável aleatória Y , foi simulado um caso numérico. Da análise dos resultados, os investigadores Bassi e Fabbris (1997) concluíram que:

- Essencialmente, os estimadores de entrevistas-reentrevistas são mais eficientes quando os desvios das respostas entre o inquérito principal e o de reentrevistas ocorrem de forma independente.
- Contudo, os estimadores de um simples inquérito podem ser mais eficientes se a variância da resposta for elevada e as correlações dos erros de resposta entre os inquéritos (principal e reentrevista) forem positivas e elevadas.
- Esta última condição é mais provável de ocorrer quando os inquiridores e supervisores forem os mesmos nas duas ocasiões. Deste modo, se os erros de resposta tendem a ser elevados e correlacionados entre os inquéritos, o uso dos resultados da reentrevista para estimar os efeitos dos inquiridores ou dos supervisores no inquérito principal não é eficiente.
- Para grandes atribuições dos inquiridores e quando forem alocados mais do que dois inquiridores por estrato, o estimador baseado num desenho amostral de única interpenetração, muitas vezes resulta em melhores estimativas do efeito dos inquiridores do que estimadores baseados num desenho amostral de dupla interpenetração para as reentrevistas.

- Do mesmo modo, sob mesmas condições no desenho amostral de única interpenetração pode ser mais eficiente do que o da dupla interpenetração para as entrevistas-reentrevistas na avaliação do erro de supervisão.
- Entretanto, há situações em que o estimador resultante de um desenho duplo é mais eficiente do que o do desenho de única interpenetração para estimar os erros de correlação.
- O estimador resultante da dupla amostragem (principal e reentrevista) é mais eficiente para combinação de valores baixos de k e m ; além disso, quanto mais baixos forem os valores de k e m mais alta é a variância simples do erro de resposta.

2.6. TAXAS DE RESPOSTA EM INQUÉRITOS POR AMOSTRAGEM

A questão das taxas de resposta em inquéritos por amostragem tem sido discutida pelos pesquisadores até os tempos de hoje. Com efeito, por exemplo, em 1981 Charlotte G. Steeh no seu artigo *“Trends in Nonresponse Rates, 1952-1979”* manifestou sua preocupação pelo facto de que as diferentes pesquisas que até então haviam sido realizadas não documentavam seus estudos sobre as crescentes taxas de não-resposta em inquéritos por amostragem. Steeh (1981) considerou dois estudos de grande dimensão na altura, realizados pelo *Survey Research Center* da Universidade de Michigan.

Esses estudos permitiram-lhe avaliar as tendências a longo prazo para as duas maiores componentes da não-resposta: as recusas e as não-entrevistas, por subgrupos urbanos usando a técnica de regressão em séries temporais. As análises feitas por Steeh demonstraram que claramente havia um crescimento substancial no total das não-respostas devido, principalmente, ao aumento das percentagens de inquiridos que se recusam a ser entrevistados. Os resultados do estudo de Steeh (1981)

demonstraram ainda que estas tendências estão relacionadas com o nível de urbanização, pois quanto mais alto for o nível de urbanização, as taxas de recusa tendem a ser mais altas – conforme ilustra a Figura 2.2 referente aos inquéritos sobre atitudes do consumidor realizados de 1954 a 1976 abrangendo cidades de três níveis diferentes.

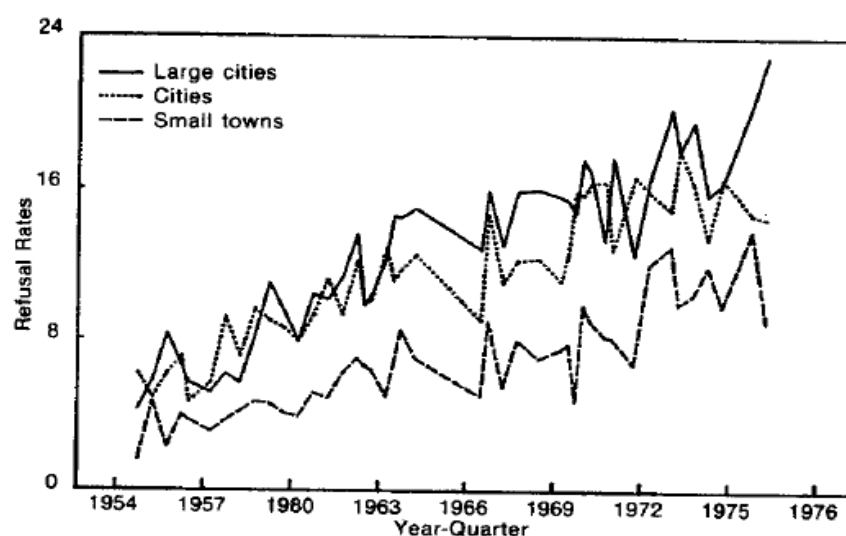


Figura 2. 2 – Taxas de Recusa nos Inquéritos sobre Atitudes do Consumidor, 1954-1976. (Steeh, 1981:45)

De acordo com Biemer e Lyberg (2003), o conceito de taxa de não-resposta é universalmente entendido e acordado entre os pesquisadores. Contudo, ainda não há uma definição universalmente aceite sobre a taxa de resposta, essencialmente por causa da elegibilidade de todas as unidades amostrais que nem sempre pode ser determinada devido a não-resposta. Este facto sugere que algum cuidado deva ser tomado ao comparar taxas de resposta de diferentes inquéritos por amostragem. Algumas medidas, diferentes das taxas de resposta, podem ser calculada usando dados provenientes de contactos com as unidades amostrais efetuados durante a recolha de dados. Tais medidas podem ser úteis para monitorar diferentes aspetos do processo de recolha de dados e, ainda servem de variáveis-chave da qualidade do processo de recolha de dados.

Com essas medidas é possível controlar o processo de recolha de dados que por sua vez pode resultar no decréscimo das taxas de não-resposta e, em muitos casos, diminuição do enviesamento da não-resposta. Os autores Biemer e Lyberg (2003) terminam sua discussão dando exemplo de variáveis-chave que podem ser calculadas e que podem apoiar no controlo do processo de recolha de dados. Exemplos dessas medidas de interesse para os pesquisadores incluem taxas de contacto, taxas de cooperação, taxas de recusa, taxas de recusa em conversão.

Como forma de exemplificar a sua abordagem, Biemer e Lyberg (2003) tomaram o conteúdo da Figura 2.3 para discutir o modelo de cálculo da taxa de resposta. Nesta figura, o total das unidades amostrais representados pela caixa (1) são repartidas em duas categorias: Unidades Resolvidas⁵ (2) e as Não-Resolvidas⁶ (3). Assim, tendo em conta que no cálculo de qualquer medida de taxa de resposta para o denominador é preciso estimar-se que unidades Não-Resolvidas de facto são fora de âmbito. Biemer e Lyberg (2003) reconhecem que vários métodos foram desenvolvidos para este assunto em função do volume e qualidade de informação disponível sobre as unidades Não-Resolvidas. Caso não haja informação disponível sobre as Não-Resolvidas, pode estimar-se (mesmo que seja grosseiramente) o número de unidades Não-Resolvidas que podem ser de âmbito do inquérito a partir da proporção das unidades Resolvidas que fazem parte da amostra. Se por exemplo, p_{Res} representa a proporção de unidades de âmbito resolvidas na população alvo e n_{NRes} denota o número de unidades Não-Resolvidas, então, o número de unidades Não-Resolvidas na população de âmbito pode ser estimado por $p_{Res} \times n_{NRes}$, recomendam os autores Biemer e Lyberg (2003).

Na óptica de Biemer e Lyberg (2003), considerando somente as unidades de âmbito, estas por sua vez podem ser classificadas em dois grandes tipos: respondentes

⁵ Unidade Resolvida é, essencialmente, a que já se determinou se a unidade é ou não de âmbito (faz parte da população alvo e, por isso, pode ser incluída na amostra) ou é fora de âmbito (não faz parte da população alvo e, por isso, não pode ser incluída na amostra).

⁶ Unidade Não-Resolvida é o caso em que não se dispõe de informação suficiente para determinar se a unidade populacional pode ser de âmbito ou não.

e as não respondentes. Ainda, as respondentes podem ser subdivididas em unidades que nunca recusam (12) e as que inicialmente recusam, mas mais tarde convertem-se em respondentes (11). As não respondentes podem ser subdivididas em três grupos: as que recusam (13), as que nunca são contactáveis (14) e as de outro tipo de não respondentes (15). Este “outro tipo” consiste de unidades que são contactadas mas não podem participar por doença, por indisponibilidade no período de entrevistas, por barreira de língua e outros motivos – Figura 2.3.

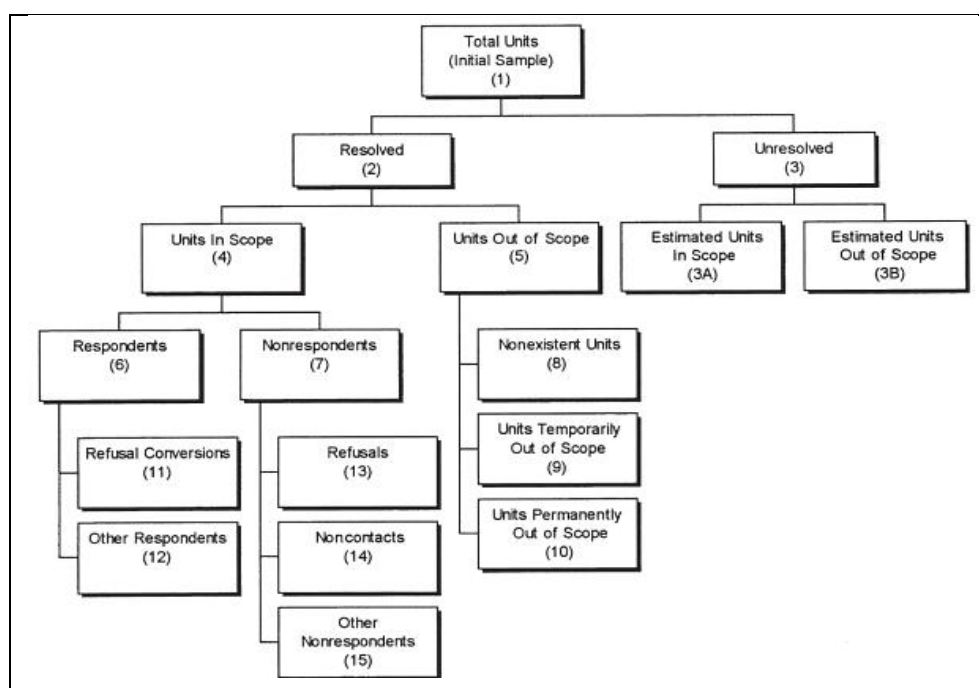


Figura 2. 3 – Disposição final da amostra com as respectivas unidades componentes de respondentes e não respondentes (Biemer e Lyberg, 2003:85).

As unidades fora de âmbito também podem ser subdivididas em: unidades não existentes (aparecem na base de sondagem, mas na realidade não existem – caixa (8)), unidades temporariamente fora de âmbito (unidades que por algum contexto constaram da base de sondagem, mas para os atuais propósitos do inquérito são considerados fora de âmbito – caixa (9)), unidades permanentemente fora do âmbito (unidades que estão fora de âmbito e não há perspectiva de serem inclusos no estudo ou nunca serão do âmbito – caixa (10)).

Nisto, Biemer e Lyberg (2003) advertem que durante o projeto do inquérito, os gestores poderão interessar-se no cálculo de distintas taxas para avaliar o progresso dos trabalhos – certamente as taxas de resposta, mas também as taxas de cooperação, taxas de barreira de língua, taxas de não-contacto, taxas de casos não-resolvidos e outras taxas.

2.6.1. Cálculo da Taxa de Resposta

Segundo Biemer e Lyberg (2003), a taxa de resposta em inquérito por amostragem é justamente o número de respondentes dividido pelo número total de unidades na amostra que fazem parte da população do âmbito do inquérito. Por outro lado, a *American Association for Public Opinion Research* (AAPOR, 2011) na sua 7ª edição do '*Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*' formaliza a definição, em termos algébricos, de seis taxas de resposta. Seja considerada a seguinte notação:

<i>RR</i>	= <i>Taxa de Resposta</i>
<i>I</i>	= <i>Entrevistas Completas</i>
<i>P</i>	= <i>Entrevista Parcial</i>
<i>R</i>	= <i>Recusa e 'break-off'</i>
<i>NC</i>	= <i>Não-Contacto</i>
<i>O</i>	= <i>Outros</i>
<i>UH</i>	= <i>Desconhecido se a unidade habitacional é ocupada</i>
<i>UO</i>	= <i>Outros casos desconhecidos</i>
<i>e</i>	= <i>Estimativa da proporção de casos de elegibilidade desconhecida que são elegíveis</i>

De acordo com AAPOR (2011), definem-se seis seguintes taxas de resposta:

- Taxa de Resposta 1 (RR1), ou a taxa de resposta mínima, é o número de entrevistas completas, dividido pelo número de entrevistas (completa mais parcial), mais o número de não-entrevistas (recusa e 'break-off' mais não-contatos e outros). Inclui ainda todos os casos de elegibilidade

desconhecida (desconhecido se a unidade habitacional é ocupada, mais outros casos desconhecidos).

$$RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UO)} \quad (2.8)$$

- Taxa de Resposta 2 (RR2) considera as entrevistas parciais como respondentes.

$$RR2 = \frac{I + P}{(I + P) + (R + NC + O) + (UH + UO)} \quad (2.9)$$

- Taxa de Resposta 3 (RR3) estima que proporção de casos de elegibilidade desconhecida é realmente elegível. A estimativa de e deve ser guiada pela melhor informação científica disponível sobre que partição de casos elegíveis pode ser feita de entre os casos desconhecidos e não se deve selecionar uma proporção a fim de aumentar a taxa de resposta. A base para essa estimativa deve ser explicitamente detalhada e indicada.

$$RR3 = \frac{I}{(I + P) + (R + NC + O) + e(UH + UO)} \quad (2.10)$$

- Taxa de Resposta 4 (RR4) atribui os casos de elegibilidade desconhecida como em RR3, mas também inclui as entrevistas parciais como entrevistados, idêntico ao que se considerou em RR2.

$$RR4 = \frac{I + P}{(I + P) + (R + NC + O) + e(UH + UO)} \quad (2.11)$$

- Taxa de Resposta 5 (RR5) é um caso especial de RR3 em que se assume que $e = 0$ (isto é, que não haja casos elegíveis entre os casos de elegibilidade desconhecida) ou no caso raro, em que não há casos de elegibilidade desconhecida. Enquanto a Taxa de Resposta 6 (RR6) faz a mesma suposição, mas inclui as entrevistas parciais como respondentes. RR5 e RR6 são apropriados apenas quando é válido assumir que nenhum dos casos desconhecidos são os elegíveis, ou quando não há casos desconhecidos. RR6 representa a taxa de resposta máxima.

$$RR5 = \frac{I}{(I + P) + (R + NC + O)} \quad (2.12)$$

$$RR6 = \frac{I + P}{(I + P) + (R + NC + O)} \quad (2.13)$$

Num inquérito por amostragem, conforme as Equações (2.8) a (2.13), o desejável é que o número de unidades com respostas completas seja igual ao total de unidades amostrais planificadas para o estudo, ou seja, ausência de quaisquer tipos de erros não amostrais. Neste caso ter-se-ia uma taxa de resposta igual a 100 por cento.

Ainda segundo Biemer e Lyberg (2003), uma das principais razões que faz com que o cálculo padronizado da taxa de resposta não seja aceite universalmente reside na dificuldade de determinar-se o valor do denominador das Equações (2.8) a (2.13). O problema surge devido a existência de casos Não-Resolvidos ou pendentes na amostra do inquérito. Ainda de acordo com Biemer e Lyberg (2003), uma das formas apropriadas do cálculo da taxa de resposta é tomar o número das unidades respondentes - caixa (6) da Figura 2.3 - dividido pelo número total das unidades amostrais de âmbito, incluindo os que são sabidos de serem da população alvo e os que se estimam pertencer à população alvo, estes, na Figura 2.3, estão representados pelas caixas (4) e (3A).

Os denominadores das Equações (2.8) a (2.13) traduzem a soma de todos os casos em que houve entrevistas completas, entrevistas iniciadas mas não completadas, recusas, não se estabeleceu contacto com a unidade amostral selecionada, outros casos em que a unidade amostral que deveria ter sido inquirida não o pôde ser (por exemplo em populações humanas, a unidade selecionada pode ter falecido, incapaz de responder por razões físicas ou psíquicas, problemas linguísticos, ou ainda outros casos reportados como erros não amostrais).

Em relação às taxas de resposta, no entanto, existe uma necessidade de um tratamento padronizado, uma vez que, quer seja para o melhor quer para o pior, esta é a medida estatística mais amplamente reportada e comparada para avaliar a qualidade de um inquérito por amostragem – defendem Biemer e Lyberg (2003). Entretanto, progressos no sentido de adotar métodos padronizados de cálculo de taxas de resposta têm sido lentos, e consequentemente, ainda é muito importante entender precisamente os métodos usados no cálculo da taxa de resposta baseado em conceptualizações.

Segundo Dey (1997); Dillman, Eltinge, Groves e Little (2002); Jones (1996) citados no artigo de Porter e Whitcomb (2003), as baixas taxas de resposta levantam a questão de quão representativos são os dados, e se os resultados de um inquérito por amostragem podem ser verdadeiramente generalizados ao universo.

2.6.2. Fatores que afetam a Taxa de Resposta

Num artigo da autoria de Deutskens, Ruyter, Wetzels e Oosterveld publicado em 2004, estes autores examinaram o efeito do tempo de *follow-ups*, os diferentes incentivos, a extensão e apresentação do questionário sobre a taxa de resposta e qualidade de resposta em ambiente *on-line* experimental. Os resultados do estudo por eles conduzido mostram que questionários curtos têm uma maior taxa de resposta, apesar de questionários longos ainda gerar uma surpreendente alta qualidade de

resposta. Além disso, estes autores concluíram que os vales postais podem ser um estímulo de resposta em questionários longos, enquanto lotarias são mais eficientes em pesquisas de curta duração.

Ainda, Deutskens et al. (2004) revelaram que estudos de *follow-up* com lotarias de prêmios pequenos, mas com maior probabilidade de ganhar são mais eficazes no aumento da taxa de resposta. Melhorar questionários com elementos visuais eleva a qualidade de resposta. Finalmente, o tempo do *follow-up* não tem nenhuma influência significativa sobre a taxa de resposta. Em pesquisas anteriores, Deutskens et al. (2004) haviam examinado um grande número de fatores que aumentam as taxas de resposta e melhoram a qualidade dos dados. Em termos de pesquisas *on-line*, os fatores mais importantes a considerar para aumento da taxa de resposta são *follow-ups*, os incentivos, a extensão e apresentação do questionário [Church (1993), Dillman (2000), Fox *et al.* (1988), Heberlein e Baumgartner (1978), Kanuk e Berenson (1975), Yammarino *et al.* (1991), Yu e Cooper (1983) citados por Deutskens *et al.* (2004)].

Ainda sobre as taxas de resposta em inquéritos por amostragem, Fan e Yan (2010) identificaram uma grande variedade de fatores que afetam as taxas de resposta em pesquisas através da internet. Eles, ainda, desenvolveram um modelo conceitual com vista a rever sistematicamente uma grande variedade de fatores que influenciam a taxa de resposta nas distintas fases da pesquisa através da internet. O estudo destes autores indica que a variedade de fatores que afetam a taxa de resposta de pesquisa na internet estão envolvidos em todas as quatro fases - apresentadas na Figura 2.4 - do processo de pesquisa através da internet.

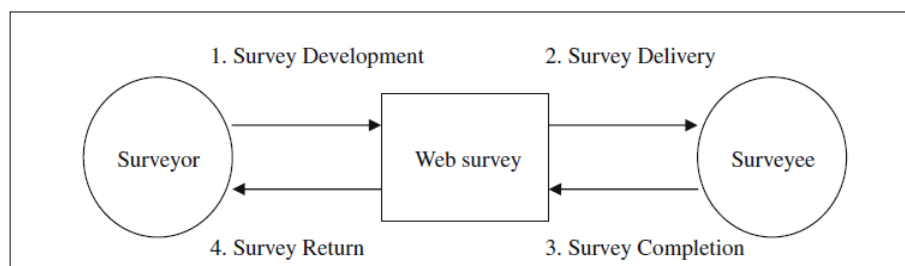


Figura 2. 4 – Processo de pesquisa através da internet (Fan e Yan, 2010:133).

Assim, segundo Fan e Yan (2010), para aumentar a taxa de resposta média, que é aproximadamente 10 por cento menor do que a por correio ou telefone, a pesquisa através da internet deve prestar atenção não só aos fatores gerais envolvidos em qualquer trabalho de pesquisa, mas também dar mais importância aos fatores específicos relacionados exclusivamente com a pesquisa da internet.

O potencial para melhorar as taxas de resposta, mudando de modo de recolha de dados para outro e as consequências para os erros de medição e de não-resposta é examinado no estudo desenvolvido por Dillman *et al.* (2009). Nesse estudo, a recolha de dados a partir de 8999 famílias foi realizada em duas fases: na fase 1 os dados foram recolhidos por entrevistas telefónicas, correio eletrónico, IVR⁷, ou internet; enquanto a fase 2 focalizou-se nos não respondentes da fase 1, e foi conduzida por telefone ou correio eletrónico. Os resultados desse estudo sugerem que a mudança para um segundo modo é um meio eficaz de melhorar a taxa de resposta. Por outro lado, esse estudo concluiu que, em geral, a mudança para um segundo modo de recolha de dados não era um meio eficaz de reduzir o erro de não-resposta com base nas variáveis demográficas.

Ainda sobre os fatores que afetam a taxa de resposta, James, Ziegenfuss, Tilburt, Harris e Beebe realizaram, recentemente em 2011, um estudo com propósito de obter colaboração de médicos em pesquisas na área de saúde. Os resultados desse estudo são encorajadores, pois revelam a eficácia que os incentivos em dinheiro pré-pago têm na otimização das taxas de resposta em pesquisas médicas. Especificamente, eles constataram que o fornecimento de incentivos imediatos para os médicos não

⁷ Resposta de Voz Interativa (IVR) é uma tecnologia de telefonia em que alguém usa um telefone *touch-tone* para interagir com um banco de dados para obter informações ou inserir dados na base de dados. A tecnologia IVR não requer interação humana sobre o telefone, pois a interação do utilizador com o banco de dados é predeterminada pelo sistema. Por exemplo, os bancos e empresas de crédito utilizam sistemas IVR para que seus clientes possam receber informações sobre suas contas imediata e facilmente sem ter que falar diretamente com uma pessoa. A tecnologia IVR é também usada para recolher informações nos inquéritos por telefone em que o utilizador é solicitado a responder a perguntas primindo as teclas de um telefone. Fonte: <http://www.webopedia.com/TERM/I/IVR.html>

respondentes de uma pesquisa nacional produziu taxas de resposta muito superiores às dos incentivos prometidos. Além disso, James *et al.* (2011) também observaram que as taxas de resposta são significativamente mais elevadas entre os que receberam incentivos imediatos em dinheiro do que os que receberam em cheques. Esta estratégia de incentivo permitiu alcançar uma taxa de resposta global que excedeu a meta de 50 por cento que havia sido estipulada.

Os resultados de James *et al.* (2011) são consistentes com as recomendações da literatura que sugerem que incentivos em dinheiro pré-pago, para populações humanas, são os mais eficazes no sentido de estimular as unidades amostrais para responder a um inquérito, de acordo com VanGeest, Johnson e Welch (2007); Flanigan, McFarlane e Cook (2008) todos citados por James *et al.* (2011).

Sabido que os erros não amostrais de cobertura, de contacto, de recusa, de resposta e outros afetam as taxas de resposta calculada com base nas Equações (2.8) a (2.13), pesquisadores da AAPOR em 2003 desenvolveram um documento com diretrizes bem claras. O principal objetivo desse documento é apoiar aos pesquisadores a calcular taxas de resposta resultantes de inquéritos via telefónica usando-se o *Random Digit Dialing* (RDD) e em inquéritos domiciliários aos agregados familiares – (Biemer e Lyberg, 2003). Com efeito, Langer (2003), no seu artigo produzido a propósito da conferência de 2003 sobre as taxas de resposta, salienta que a própria AAPOR reconhece a dificuldade que se experimenta no cálculo de taxas de resposta. Em 2003, a AAPOR emitiu um padrão de normas e definições referentes à disposição do plano amostral sobre o qual as taxas de resposta podem ser calculadas. Esse padrão de definições é útil, mas não constitui o fim da complexidade da questão do cálculo de taxas de resposta. Por outro lado, essas normas ainda não foram promulgadas para todos os vários desenhos amostrais, gestão de amostras e procedimentos de seleção de respondentes que pode afetar as taxas de resposta. Ainda no artigo de Langer (2003), a AAPOR adverte que "*As comparações de taxas de resposta entre inquéritos com desenhos amostrais diferentes não são possíveis*". Dependendo da natureza dos erros não amostrais envolvidos, os padrões definidos

pela AAPOR em si são complexos na medida em que oferece quatro métodos diferentes de cálculo de taxas de cooperação, três métodos para cálculo de taxas de recusa, três para as taxas de contacto, e seis para as taxas de resposta global. Estas últimas são apresentadas nesta dissertação pelas Equações (2.8) a (2.13).

2.6.3. O Enviesamento devido à Não-Resposta

Segundo Groves e Peytcheva (2008), muitas pesquisas por amostragem seguem o paradigma de inferência que assume taxas de resposta de 100 por cento para um desenho amostral probabilístico. Ou seja, o enviesamento das estimativas e a medida de seus erros padrão permitem declarações probabilísticas sobre as características da população, quando todos os elementos da amostra são observados. Entretanto, quando uma parte dos elementos da amostra não é observada em consequência das não-respostas, algumas propriedades da amostragem probabilística sobre inferência não são efetivas, a menos que algum modelo do impacto do erro de não-resposta seja considerado. Os pesquisadores passam por desafios a este paradigma básico de inferência por causa da queda das taxas de resposta em inquéritos por amostragem em todo os países mais ricos do mundo [de Leeuw e de Heer (2002) citados por Groves e Peytcheva (2008)]. Os desafios são agravados pelo facto de que os projetos de pesquisa que buscam altas taxas de resposta enfrentam custos crescentes, gerados por esforços repetidos para obter acesso às unidades amostrais e para solucionar quaisquer dúvidas das pessoas da amostra.

Segundo Groves (2006), a não-resposta em inquéritos aos AF's está a crescer inflacionando os custos de pesquisas na tentativa de atingir elevadas taxas de resposta. No entanto, o enviesamento da não-resposta em estimativas é apenas indiretamente relacionado com a taxa de não-resposta. Um parâmetro chave que determina a relação entre a taxa de não-resposta e o enviesamento da não-resposta é o quão correlacionadas estão a variável de interesse da pesquisa e a propensão de resposta.

A este respeito, Groves e Peytcheva (2008) fazem referência de um estudo realizado em 2006 por Groves no qual examinou um conjunto de 30 estudos para estimar o enviesamento da não-resposta de estatísticas descritivas. Nesse estudo, Groves analisou o impacto da não-resposta no enviesamento das estimativas em inquéritos domiciliários. A principal motivação de Groves prendeu-se com o facto de nos últimos tempos muitos inquéritos às famílias nos EUA estarem a experimentar elevadas taxas de recusa. Segundo Groves (2006), a não-resposta pode, mas não necessariamente, induzir enviesamento da não-resposta nas estimativas de um inquérito. Resultados empíricos da análise de Groves ilustraram casos em que a ligação entre as taxas de não-resposta e o enviesamento de não-resposta não é notória. Apesar disso, os padrões de profissionalismo continuam a exigir dos pesquisadores elevadas taxas de resposta. As sínteses de pesquisas que visam estimar o enviesamento da não-resposta mostram que, muitas das vezes, o referido enviesamento existe. Na óptica de Groves, a pergunta lógica que se coloca é que vantagens têm os inquéritos por amostragem probabilística sabido que estes experimentam taxas de não-resposta elevadas.

Da análise do estudo de Groves (2006), ele acha que a taxa de não-resposta, por si só, é um preditor pobre da magnitude do enviesamento em 319 diferentes estimativas que foram calculadas a partir desses 30 estudos. Esta conclusão de Groves prende-se com o facto de ele ter constatado que a variação das taxas de não-resposta explica em apenas cerca de 11 por cento da variação do enviesamento da não-resposta nas diferentes estimativas de interesse.

Ainda no mesmo estudo realizado por Groves em 2006, ele cita os pesquisadores Curtin, Presser, Singer (2000), Keeter *et al.* (2000), Merkle e Edelman (2002) que referem que as alterações nas taxas de não-resposta não alteram necessariamente as estimativas da pesquisa. No entanto, a prescrição mais comum para os pesquisadores é minimizar as taxas de não-resposta. Por exemplo, Groves cita Alreck e Settle (1995) que diz ser óbvio que seja importante fazer o máximo possível para reduzir a não-resposta e incentivar a uma taxa de resposta adequada. Uma taxa adequada significa

para Babbie (2007) citado por Groves o seguinte: a taxa de resposta de pelo menos 50% é considerada adequada para análise e relatório; a taxa de resposta de 60% é boa; uma taxa de resposta de 70% é muito boa. Outra advertência citada por Groves é de Singleton e Straits (2005) que recomenda que é muito importante prestar atenção às taxas de resposta, pois para inquéritos por entrevista uma taxa de resposta de 85% é minimamente adequada e uma abaixo de 70% há uma probabilidade elevada de um enviesamento das estimativas.

Assim, para Groves (2006), não há taxa de resposta mínima abaixo da qual as estimativas da pesquisa estão necessariamente sujeitas ao enviesamento.

Entretanto, de acordo com Wagner (2010) a taxa de resposta tem sido um indicador tradicional de qualidade dos dados de um inquérito por amostragem. No entanto, uma pesquisa empírica recente classificou a taxa de resposta como sendo um indicador *proxy* do enviesamento da não-resposta. Por isso, Wagner (2010) propôs uma nova medida para o risco da não-resposta, a que chamou de *Fraction of Missing Information* (FMI)⁸. A FMI é proposta como uma alternativa para a taxa de resposta e foi desenvolvida como parte dos métodos de tratamento de dados em falta, mede o nível de incerteza sobre os valores que seriam de imputar para os atuais não respondentes. Desenvolvida como uma ferramenta para monitorar o decurso da recolha de dados, a FMI pode ajudar aos pesquisadores a maximizar a informação no conjunto de dados finais e, ainda a FMI é aplicada em dados de pesquisas por telefone e cara-a-cara.

Ainda segundo Wagner (2010), o método mais simples para estimar a FMI é a criação de imputações múltiplas sob um modelo pré-assumido. Um conjunto de dados para imputação de alta qualidade produz menor grau de incerteza sobre os valores imputados, ou seja, a qualidade de dados que se têm melhora as estimativas por permitir imputar as não-respostas com maior rigor.

⁸A FMI pode ser entendida como 'Fração de Informação Omissa' – tradução livre do autor deste trabalho.

Em 1987, Rubin citado por Wagner (2010), mostrou que a incerteza sobre valores imputados pode ser caracterizada por duas componentes: a variância inter-imputação e a média da variância intra-imputação. A média da variância intra-imputação é a componente de variância devido ao erro de amostragem. Esta corresponde à variância habitualmente estimada para uma estatística baseada em dados completos, mas tomada a média devido aos múltiplos conjuntos de dados imputados, e assume-se que os valores omissos são conhecidos.

A variância inter-imputação é a componente que resulta da criação de uma distribuição aproximada de valores imputados. Esta, por sua vez, é a variância das estimativas resultantes de cada um dos conjuntos de dados imputados. Quanto maior for a diferença entre essas estimativas, maior se torna esta componente. A componente de variância inter-imputação geralmente diminui quando dados completos incluem preditores da variável que tem valores omissos (Wagner, 2010).

Estas duas componentes da variância são combinadas⁹ e incorporam um determinado ajuste de modo a formar uma estatística a que Wagner chamou de 'Variabilidade total' associada à média do parâmetro de interesse. Por fim, a quantidade utilizada para avaliar o impacto da não-resposta é a FMI definida como o rácio entre a Variabilidade inter-imputação e a Variância total das estimativas do inquérito.

Para Wagner (2010), se conceptuarmos a amostra como uma matriz de dados, onde os valores em falta representam os não respondentes, então monitorar a variância inter-imputação parece uma escolha natural para avaliar a incerteza sobre a não-resposta. No contexto da não-resposta nas pesquisas, a FMI é uma medida de incerteza sobre os valores imputados para os não respondentes. Por exemplo, numa imputação utilizando outros dados auxiliares que se correlacionam com as variáveis de pesquisa (i.é, dados reais recolhidos de respondentes) a FMI vai ser reduzida e muito

⁹ Mais detalhes sobre a definição da FMI e equações de cálculo são apresentados no artigo de Wagner (2010:241)

abaixo da taxa de não-resposta. Neste caso, pode-se ter uma melhor noção da qualidade dos dados obtidos após os ajustes da não-resposta feitos com recurso aos dados da base de sondagem e de outra informação auxiliar.

3. METODOLOGIA DO ESTUDO

Neste capítulo é apresentada a metodologia seguida para elaboração desta dissertação. Na Secção 3.1 são apresentados aspetos metodológicos gerais que inclui as consultas feitas às fontes consideradas relevantes para o tema na revisão da literatura. A metodologia desta dissertação inclui análise dos dados de três inquéritos nacionais realizados pelo INE-M. Assim, mais adiante nas Secções 3.2 a 3.4, são apresentados, resumidamente, os aspetos metodológicos e procedimentos aplicados na condução desses três inquéritos. A apresentação da metodologia dos três inquéritos visa identificar as semelhanças ou diferenças de procedimentos empregues, que medidas foram tomadas pelo INE-M para prevenir e ou controlar os erros não amostrais. No fim deste capítulo (Secção 3.5) são apresentadas as técnicas empregues para análise de taxas de resposta dos três inquéritos e os fatores de não-resposta.

3.1. ASPETOS GERAIS DA METODOLOGIA USADA

Conforme se referiu anteriormente, a presente dissertação foi elaborada com base na pesquisa bibliográfica em literatura física e de artigos científicos considerados relevantes para o tema. A pesquisa de informação disponível serviu para fazer um enquadramento teórico sobre as abordagens do tema por distintos autores. As palavras-chave usadas na pesquisa bibliográficas são as ligadas às medidas de qualidade em inquéritos por amostragem ou ainda aos autores que se debruçaram nestas matérias. A pesquisa de informação foi feita de forma sistemática através das bibliotecas físicas e recursos on-line da b-On, Wiley-Blackwell, ISI Web of Knowledge, JSTOR. Por outro lado, a revisão da literatura serviu para identificar e descrever os tipos de erros não amostrais que ocorrem nas distintas fases de um inquérito.

Com este estudo também pretende-se avaliar o quão variam as taxas de respostas nos inquéritos por amostragem como resultado de ocorrência de erros de não-resposta durante o processo do inquérito. O intuito de se identificar os principais

fatores que influenciam os erros de não-resposta é, principalmente, o de avaliar as suas consequências nas taxas de resposta, para que se possam tomar medidas com vista ao seu controlo e prevenção de modo que a taxa de resposta de determinado inquérito não possa ser afetada significativamente.

Com base nos relatórios de procedimentos de amostragem elaborados para condução de cada um dos inquéritos referidos no 4º parágrafo do Capítulo 1, com recurso às bases de dados (BD's) dos inquéritos, vão ser identificadas variáveis ligadas aos resultados das entrevistas dos Agregados Familiares (AF's)¹⁰ de onde são entrevistados os respetivos chefes na Cidade de Maputo¹¹.

Mais adiante, no Capítulo 4 referente à apresentação e análise dos dados, vão ser usadas três BD's correspondentes aos três inquéritos ora mencionados. Recorrendo-se ao pacote estatístico SPSS¹² versão 18 vão ser elaboradas tabelas de frequências dos respondentes dos três inquéritos. Identicamente, vão ser obtidas tabelas de frequências das variáveis 'Resultados da entrevista' e 'Número de visitas ao AF'. Com base nas frequências, usando-se as Equações (2.8) a (2.13), serão calculadas taxas de respostas associadas a cada inquérito.

¹⁰ Define-se Agregado Familiar como sendo todo o grupo de pessoas ligadas ou não por laços de parentesco, que vivem na mesma casa (habitação) e compartilham a maior parte das despesas de alojamento e alimentação (INE-M, Censo 2007).

¹¹ Chefe do Agregado Familiar é a pessoa responsável pelo agregado ou que, para efeitos do censo ou inquérito, é indicada como tal pelos restantes membros. Em cada AF deverá haver sempre um chefe e este deve ser uma pessoa aí residente, podendo estar presente ou não no momento da entrevista, desde que a ausência seja inferior a 6 meses (INE-M, Censo 2007).

¹² **SPSS** é um software aplicativo (programa de computador) do tipo científico, acrónimo de *Statistical Package for the Social Sciences* - pacote estatístico para as ciências sociais. Este pacote de apoio a tomada de decisão inclui aplicação analítica, Data Mining, Text Mining e estatística que transformam os dados em informações importantes. Fonte: <http://pt.wikipedia.org/wiki/SPSS>

3.2. PROCEDIMENTOS METODOLÓGICOS DO IDS2003

Os procedimentos metodológicos que a seguir se apresentam nesta secção são parte integrante dos documentos 'Recomendações do desenho da Amostra do IDS2003' e 'Manual do Inquiridor' ambos elaborados pelo INE-M.

O INE-M em colaboração com o Ministério da Saúde realizaram o Inquérito Demográfico e de Saúde em 2003 com a assistência técnica da *Macro International* e apoio financeiro da USAID.

O IDS2003 foi uma pesquisa junto aos AF's, por amostragem, que recolheu informações sobre características demográficas e aspectos ligados à saúde da população moçambicana. Este tipo de pesquisa faz parte de um programa mundial de inquéritos. Em Moçambique, este foi o segundo inquérito a ser realizado, depois do primeiro em 1996.

Este inquérito recolheu também dados sobre traumatismo, planeamento familiar, saúde materno-infantil, nutrição, imunização, saúde reprodutiva, VIH/SIDA. Foram entrevistadas mulheres dos 15 aos 49 anos e homens dos 15 aos 64 anos, selecionados aleatoriamente em todo o país.

3.2.1. Objetivos do IDS2003

Em termos gerais, os objetivos do IDS2003 foram: (i) Fornecer informação estatística para a formulação de políticas e implementação de programas sobre população, saúde e para o monitoramento de programas e planificação; (ii) Aumentar e atualizar a base internacional de dados sobre população e saúde para acompanhamento e avaliação; (iii) Atualizar as taxas demográficas, particularmente as taxas de fecundidade e de mortalidade materno-infantil; (iv) Ter informação sobre uso de métodos

contraceptivos; (v) Obter informação, comportamentos, atitude e práticas sobre VIH/SIDA.

3.2.2. A Amostragem do IDS2003

A amostra do IDS2003 foi probabilística e representativa, desenhada principalmente para produzir estimativas precisas de taxas demográficas, particularmente de indicadores de saúde materno-infantil e de conhecimento e uso de métodos contraceptivos para o país todo e para os estratos urbano e rural do país, e ainda separadamente, para as 11 províncias incluindo a Cidade de Maputo.

A base de sondagem do IDS2003 foi a lista derivada do Inquérito aos Agregados Familiares sobre despesas e orçamento familiar (IAF) realizado pelo INE-M entre 2002 e 2003. Assim, a amostra do IDS2003 foi uma sub amostra do IAF2002¹³.

A seleção da amostra do IDS2003 foi feita em três etapas: na primeira etapa foram selecionadas 604 Unidades Primárias de Amostragem (UPA's)¹⁴ com probabilidade proporcional à dimensão das UPA's em termos de AF's; na segunda, foi selecionado igual número de áreas de enumeração (AE's) com probabilidades iguais dentro de cada UPA amostral. Finalmente, na terceira etapa, em cada AE foram selecionados AF's para entrevistas. Os AF's foram selecionados com probabilidades iguais.

¹³ A amostra do IAF foi extraída da Amostra Mãe. A Amostra Mãe foi elaborada pelo INE-M com base nos resultados do Censo de 1997.

¹⁴ A UPA (Unidade Primária de Amostragem) é um conjunto de áreas de enumeração adjacentes, que contêm entre 400 e 600 agregados na zona urbana, e entre 400 e 500 agregados na zona rural. A Unidade Secundária de Amostragem (USA) é a área de enumeração delimitada durante a cartografia censitária de 1997, que contêm entre 120 e 150 agregados na zona urbana, e entre 80 e 100 agregados na zona rural.

Quanto à dimensão da amostra do IDS2003 e sua repartição, referir que o IDS2003¹⁵ foi desenhado para ser representativo por cada domínio mais baixo (província) e incluiu 52 conglomerados (AE's) por província com a excepção das províncias de Nampula e Zambézia com 68 AE's cada, devido ao peso das suas populações no total do país. O Quadro 3.1 mostra distribuição de 14.496 AF's esperados por província correspondendo a 12.700 mulheres elegíveis¹⁶ distribuídos por área de residência urbano e rural. Esta dimensão foi determinada tendo em conta a taxa de resposta mínima requerida de 90 por cento nos AF's.

Província	Agregados Familiares esperados			Mulheres Elegíveis esperadas		
	Total	Urbano	Rural	Total	Urbano	Rural
Niassa	1.248	600	648	1.100	529	571
Cabo Delgado	1.248	192	1.056	1.100	169	931
Nampula	1.632	384	1.248	1.300	306	994
Zambezia	1.632	240	1.392	1.300	191	1.109
Tete	1.248	288	960	1.100	254	846
Manica	1.248	600	648	1.100	529	571
Sofala	1.248	480	768	1.200	462	738
Inhambane	1.248	288	960	1.100	254	846
Gaza	1.248	456	792	1.100	402	698
Maputo Província	1.248	720	528	1.100	635	465
Maputo Cidade	1.248	1.248	0	1.200	1.200	0
Moçambique	14.496	5.496	9.000	12.700	4.930	7.770

Quadro 3. 1 - Dimensão e repartição da Amostra do IDS2003. Fonte: INE-M, IDS2003.

Por província esta amostra forneceu estimativas com 95 por cento de confiança. A amostra é independente em cada província, contudo a dimensão da amostra não permite comparações entre as populações urbana e rural de cada província.

¹⁵ O cálculo da dimensão e sua repartição foi determinado por um consultor de amostragem da *Macro International*, em Maio de 2002.

¹⁶ Um elegível ao IDS2003 é toda a pessoa cuja idade é dos 15 aos 49 anos (mulheres) ou dos 15 aos 64 anos (homens).

Província	Agregados Familiares esperados			AF por AE	Conglomerados (AE's)		
	Total	Urbano	Rural		Total	Urbano	Rural
Niassa	1.248	600	648	24	52	25	27
Cabo Delgado	1.248	192	1.056	24	52	8	44
Nampula	1.632	384	1.248	24	68	16	52
Zambezia	1.632	240	1.392	24	68	10	58
Tete	1.248	288	960	24	52	12	40
Manica	1.248	600	648	24	52	25	27
Sofala	1.248	480	768	24	52	20	32
Inhambane	1.248	288	960	24	52	12	40
Gaza	1.248	456	792	24	52	19	33
Maputo Provincia	1.248	720	528	24	52	30	22
Maputo Cidade	1.248	1.248	0	24	52	52	0
Moçambique	14.496	5.496	9.000	24	604	229	375

Quadro 3. 2 - Dimensão e repartição da Amostra do IDS2003. Fonte: INE-M, IDS2003

A amostra do IDS2003 permitiu obter estimativas fiáveis para os domínios Nacional, Urbano e Rural, Regional¹⁷ e Provincial. Ainda, esta amostra contou somente com os AF's residentes em habitações particulares. Portanto, foram excluídos os agregados familiares, as mulheres elegíveis residentes em estabelecimentos coletivos como hotéis, penitenciárias, quartéis, lares de estudantes os quais com outros residentes naquelas condições, juntos representaram menos de meio por cento do total da população moçambicana (INE-M, Censo 1997).

3.2.3. Listagem de Agregados familiares do IDS2003

Para obter uma amostra atualizada de agregados da Amostra Mãe, a seleção de agregados no último estágio de amostragem foi baseada numa listagem recente dentro de cada AE amostral. Para evitar tendências de uma seleção de agregados feita pelo inquiridor, recomendou-se fazer a listagem de agregados numa AE amostral pelo

¹⁷O território moçambicano é dividido administrativamente em três regiões: Norte abrangendo as províncias de Niassa, Cabo Delgado e Nampula; Centro que abrange as províncias de Zambézia, Tete, Manica e Sofala. A região Sul abrange as províncias de Inhambane, Gaza e Maputo (inclui a Cidade de Maputo – capital do país).

menos um mês antes da recolha de dados. Assim, a seleção de agregados seria melhor controlada no escritório central ou provincial.

O Agente de Listagem (AL)¹⁸ levava consigo ao campo um mapa ou croquis¹⁹ da AE. Num mapa ou Croquis constam as referências geográficas essenciais e ou infraestruturas como escolas, igrejas, ruas, caminhos, rios, lagoas, estradas – conforme se pode ver o modelo no Anexo A2 - com as quais devia estabelecer uma correspondência das referências geográficas observáveis no terreno. O AL utilizou uma Ficha de Listagem (ver modelo no Anexo A3) para registar todos os AF's encontrados na AE.

O AL devia ter o cuidado de não omitir agregados familiares cujas habitações estivessem escondidas. Em algumas zonas, as residências dos AF's estão construídas de forma desordenada, por isso que podem ser facilmente omitidas.

3.2.4. Questionários do IDS2003

No IDS2003 foram usados três tipos de questionários: (i) Questionário do Agregado Familiar, (ii) Questionário das Mulheres e (iii) Questionário dos Homens.

Cada AF selecionado foi visitado e entrevistado. Para o Questionário do Agregado Familiar a entrevista foi feita ao respetivo chefe ou a um outro membro que pudesse responder pelo chefe e ajudasse a identificar todos os que residem habitualmente no AF e visitantes que tivessem passado noite anterior à entrevista no AF, o que permitiu apurar as pessoas elegíveis dentro deste. Ainda neste Questionário recolheu-se informação básica acerca de cada pessoa, o sexo, a idade e o nível de

¹⁸ Um Agente de Listagem é mais um membro da brigada de campo cujas funções principais consistem em delimitar as unidades geográficas selecionadas e elaborar a lista completa de todos os agregados familiares residentes no seu interior.

¹⁹ Mapa ou croquis da AE é uma representação topográfica da área de enumeração que inclui principais elementos físicos, tais como montanhas, vias de acesso, infraestruturas sociais, rios, ruas, níveis de altitudes.

instrução. Esta informação foi usada para calcular taxas demográficas e para avaliar a qualidade da implementação da amostra. Através do Questionário do Agregado Familiar o inquiridor identificou as pessoas que viriam a ser entrevistadas nos questionários individuais.

Nos questionários individuais eram entrevistadas as mulheres dos 15 aos 49 anos e os homens dos 15 aos 64 (somente nos agregados selecionados para entrevistas de homens).

Importa referir que tanto os membros (pessoas habitualmente residentes no AF) quanto aos visitantes (pessoas que não residem habitualmente no agregado, mas aí passaram a noite anterior à entrevista) eram entrevistados nos questionários individuais.

3.2.5. Organização do Trabalho de campo do IDS2003

Em cada AE amostral era necessário entrevistar pelo menos 18 mulheres elegíveis e, para o efeito foram visitados 24 AF's. Para efeitos de compensação, todas as mulheres elegíveis num AF eram entrevistadas, dado que poderiam ser encontrados AF's sem mulheres elegíveis. Na amostra de homens eram necessários 6 a 7 entrevistas por AE amostral. Este facto implicou que o inquiridor tivesse de entrevistar pelo menos um homem elegível em cada quatro AF's selecionados.

Cada província teve uma brigada para a recolha de dados. Cada brigada, excetuando Zambézia e Nampula, teve a seguinte composição: um controlador, quatro inquiridoras, um inquiridor, um motorista. As províncias de Zambézia e Nampula, devido ao tamanho das respetivas amostras tiveram duas brigadas cada com a mesma composição ora descrita.

Antes do trabalho de campo, o pessoal envolvido no IDS2003 em todas as fases beneficiou de uma formação. A formação foi conduzida por membros da Direcção Central do IDS2003 e compreendeu sessões teóricas sobre a condução da entrevista, entrevistas simuladas na sala e sessões de prática no terreno.

3.2.6. Procedimentos aplicados na Condução de entrevistas do IDS2003

A entrevista é uma técnica de recolha de dados e/ou de obtenção de informação através de perguntas efetuadas à pessoas idóneas para sua resposta imediata e direta. Efetuar uma entrevista com êxito é uma arte e como tal não deve ser tratada como um processo mecânico. Deve ser conduzida como uma conversa normal entre duas (ou mais) pessoas, o que implica a observância de regras básicas para o seu êxito.

Alguns aspetos importantes que se tinham em conta durante a entrevista do IDS2003:

1) Acesso ao entrevistado

O inquiridor e o entrevistado não se conheciam. Por esta razão, a primeira impressão da aparência do inquiridor, as suas primeiras ações e as palavras expressas foram de vital importância para ganhar a cooperação do entrevistado. Uma vez que se encontrava em presença do entrevistado, a primeira coisa que o inquiridor devia fazer foi apresentar-se amavelmente, indicando o seu nome pessoal, o nome da instituição para a qual trabalha, mostrar o seu crachá e explicar o que desejava da entrevista.

Era importante conseguir um contacto inicial positivo. Não era conveniente usar perguntas como: *"está muito ocupado?"*. *"Pode conceder-me alguns minutos?"* ou *"poderia responder-me algumas perguntas?"*.

Igualmente, era recomendado utilizar uma fórmula que convidasse à aceitação *"eu gostaria de fazer-lhe algumas perguntas..."*. Por outro lado, era também muito

importante que o inquiridor desse a conhecer aos entrevistados os objetivos do inquérito antes de desenvolver o questionário. Como no questionário previa-se a realização de perguntas sobre os membros do agregado, foi importante que os referidos objetivos fossem também do conhecimento dos entrevistados.

Caso o inquiridor estivesse acompanhado do supervisor ou outro membro da brigada de trabalho, devia apresentá-lo no início da entrevista. As explicações jogam um papel muito importante na vontade das pessoas para responder às perguntas.

2) Confidencialidade das respostas

Antes de fazer a primeira pergunta era necessário dar a conhecer o carácter confidencial da informação a todos os membros presentes do AF. O inquiridor devia explicar que *“não se publicaria nomes das pessoas em nenhum caso; que a informação seria publicada de forma agregada”*. Por nenhum motivo era permitido mostrar questionários preenchidos à pessoas alheias ao IDS2003.

3) Neutralidade

Os questionários tinham sido cuidadosamente desenhado para evitar a possibilidade de sugerir respostas ao entrevistado, portanto, tornava-se rigorosamente importante que o inquiridor se mantivesse neutro em relação ao conteúdo dos questionários do IDS2003.

Quando o entrevistado respondesse de maneira vaga ou imprecisa, o inquiridor devia indagar de maneira neutra dizendo *“pode explicar melhor?”*, *“não pude perceber o que disse, poderia repetir?”*, *“Não há pressa, tem todo o tempo para pensar”*. Por nenhum motivo o inquiridor devia interpretar o expressado pelo entrevistado.

Nunca se podia fazer notar, quer que fosse pela expressão do rosto, quer pelo tom da voz que o entrevistado deu uma resposta incorreta, imprecisa ou errada.

Algumas vezes o entrevistado perguntava ao inquiridor sobre sua opinião ou ponto de vista. O inquiridor devia sugerir que "*Sua opinião é a que tem valor para o inquérito*" mas que depois da entrevista poderia dedicar-lhe alguns minutos para conversar, se assim o desejasse.

Ainda, caso o entrevistado vacilasse em responder alguma pergunta ou negasse fazê-lo, o inquiridor devia tratar de vencer essa resistência, explicando uma vez mais, a natureza confidencial da informação e que no IDS2003 participam pessoas de todo o país. Se apesar disso, o entrevistado continuasse a negar responder, o inquiridor devia comunicar imediatamente ao supervisor. Uma vez feitas todas as perguntas, o inquiridor devia solicitar, cortesmente, a informação em falta.

4) Controle da entrevista

O inquiridor é quem dirigia a entrevista; por isso devia conduzi-la condignamente. Quando se pusesse em dúvida a autoridade do inquiridor, para fazer certas perguntas, era conveniente explicar ao entrevistado que ele foi treinado para essa tarefa e que seu trabalho consiste em fazer perguntas dessa natureza.

Se o entrevistado desse respostas de temas alheios ou falasse de assuntos que não tivessem nada a ver com o IDS2003 não era-lhe interrompido, mas na primeira oportunidade, com muita criatividade, o inquiridor tinha de fazer de novo a pergunta. Era crucial e necessário manter um bom ambiente durante a entrevista. Quando o entrevistado notasse que o inquiridor era amável, simpático e desinibido, ele estaria mais inclinado a responder às perguntas do IDS2003 sem reparos.

5) Lidando com pessoas indecisas

Em muitas ocasiões o entrevistado respondia "*não sei*", dava uma resposta com evasivas, traduzia o que tivesse dito anteriormente ou recusava responder às perguntas. Nestes casos, o inquiridor tratava de dar-lhe mais confiança e fazer com que ele se sentisse mais cómodo, antes de continuar com a pergunta seguinte.

6) Língua da entrevista

As perguntas dos questionários do IDS2003 poderiam ser traduzidas para a língua local. Era muito importante não alterar o significado das perguntas quando tivesse que usar suas próprias palavras na tradução para outras línguas.

7) Fim da entrevista

Uma vez finalizada a entrevista, era recomendado ao inquiridor que revisse o questionário na presença do entrevistado. E, o inquiridor antes de se retirar da habitação do agregado tinha de agradecer a colaboração prestada e se despedia.

3.2.7. Procedimento para Contactos aos agregados familiares

Localização dos agregados selecionados

No âmbito do IAF2002/03 foi feita listagem dos agregados familiares nas áreas de enumeração selecionadas. Essas listas foram utilizadas para a selecção dos agregados familiares que seriam entrevistados no IDS2003. Para localizar os agregados familiares selecionados o inquiridor contou com ajuda de guias locais²⁰.

Problemas no contacto com um agregado familiar

Em alguns casos havia dificuldades em identificar AF's selecionados porque as pessoas mudaram-se para outro lugar, ou as listas continham nomes mal redigidos. Eis alguns exemplos de problemas que ocorriam e algumas sugestões de como eram resolvidos:

- (a) **O agregado familiar mudou de residência e esta está vaga.** Se o agregado familiar mudou de residência de onde foi listado e ninguém vive nela, devia

²⁰ Guia local é uma pessoa selecionada para tal, de preferência autoridade administrativa ou tradicional da AE. O guia local deve conhecer com perfeição as vias de acesso mais seguras e rápidas para os AF's, por isso que apoia na identificação dos limites da AE e de habitações dos AF's. O guia deve ser conhecido pela maioria dos membros da comunidade local e exercer influência sobre eles.

considerar-se casa desocupada e registar o código correspondente na capa do Questionário do Agregado Familiar.

- (b) **O agregado familiar listado mudou de residência, mas há novos moradores.** Neste caso, entrevistavam-se os novos moradores.
- (c) **O Chefe do agregado familiar mudou.** Em alguns casos, a pessoa listada como chefe do agregado pode ter-se mudado ou morrido. Entrevistava-se o novo chefe do agregado familiar que reside no endereço listado.
- (d) **A habitação está fechada e os vizinhos dizem que os moradores viajaram e regressarão dias depois.** Este agregado devia ser revisitado mais tarde e pelo menos duas vezes para ver se eles regressaram. Caso contrário, o inquiridor registava na capa do questionário “Todo o AF ausente”.
- (e) **A habitação do agregado selecionado não foi encontrada na AE porque foi destruída.** O inquiridor registava na capa do questionário “Casa destruída”.
- (f) **Ausência de pessoas idóneas no agregado ou outra situação imprevista.** O inquiridor regressava novamente mais tarde (ou no dia seguinte). Caso não suprisse a dificuldade depois de esgotadas as visitas, o inquiridor registava na capa do questionário “Outro”.

Identificação de pessoas elegíveis

No IDS2003 utilizou-se o Questionário do Agregado Familiar para identificar as pessoas elegíveis para o Questionário Individual das mulheres ou dos homens. Os exemplos que se seguem mostram os procedimentos metodológicos que se seguiram consoante o cenário tido na identificação das pessoas elegíveis e para conseguir uma entrevista.

Falta de pessoas elegíveis. Em alguns agregados familiares, podia não haver pessoas elegíveis. Nestes agregados, devia preencher-se o Questionário do Agregado Familiar e não era acompanhado de nenhum Questionário Individual.

A pessoa elegível está ausente. Caso a pessoa elegível estivesse ausente da casa no momento da entrevista deveria contactar mais tarde a residência, pelo menos duas vezes, procurando remarcar as visitas em diferentes momentos do dia. Caso não suprisse a dificuldade depois de esgotadas as visitas, o inquiridor registava na capa do questionário individual “Ausente”.

A pessoa Recusa-se ser entrevistada. A disponibilidade e a vontade da pessoa em ser entrevistada dependem em larga medida das impressões iniciais que se tiver do inquiridor. Ao inquiridor era-lhe exigido que se apresentasse, explicasse os objetivos da sua visita, e dissesse a duração da entrevista. Caso fosse necessário, dizia-lhe que a informação que prestar será considerada confidencial. Se a pessoa ainda continuasse a resistir à entrevista, podia ser que não quisesse naquele momento. Agendava-se outra ocasião para a entrevista. Se ela continuasse a recusar, então escrevia na capa do questionário individual “Recusa total” e informava-se o supervisor de campo.

Recusa Durante a Entrevista. O entrevistado podia interromper a entrevista por alguma razão ou porque não pretendesse responder a todas as perguntas naquele momento. Se por qualquer motivo a entrevista tivesse que ser interrompida marcava-se outra ocasião para completá-la. Caso não suprisse a dificuldade depois de esgotadas as visitas, o inquiridor registava na capa do questionário individual “Recusa durante a entrevista” e participava o sucedido ao supervisor.

Informante Incapaz. Havia casos em que não podia entrevistar-se uma pessoa por doença ou era incapaz de compreender as perguntas, ou porque ela é

surda. Nestes casos, registava-se na capa do questionário individual “Incapacitado”.

3.3. PROCEDIMENTOS METODOLÓGICOS DO IFTRAB2004

Os procedimentos metodológicos que se apresentam nesta secção são derivados principalmente do Manual do Inquiridor do IFTRAB2004 elaborado por INE-M em 2004.

O INE-M conduziu o IFTRAB2004 durante 12 meses, desde Julho de 2004. Este inquérito foi financiado por um Projecto Escandinavo. A USAID contribuiu na assistência técnica para o desenho da amostra.

A base de amostragem para o IFTRAB2004 foi a Amostra Mãe derivada dos dados e cartografia do Censo de 1997. A metodologia usada para o desenho da Amostra Mãe está documentada no relatório “Desenho da Amostra Mãe Derivada do Recenseamento Populacional de 1997”. Na elaboração do relatório metodológico do IFTRAB2004, um consultor fez referência à documentação da Amostra Mãe e vários relatórios de inquéritos prévios conduzidos pelo INE-M. Os dados do IAF2002/03 foram usados para estimar o nível de precisão que se podia esperar dos resultados mais importantes do IFTRAB2004.

3.3.1. Objetivos do IFTRAB2004

O objetivo principal do IFTRAB2004 foi medir as características da força de trabalho em Moçambique, inclusive indicadores de população economicamente ativa, taxas de desemprego e subemprego, como também a natureza do trabalho infantil. O INE-M teve colaboração do Ministério do Trabalho, UNICEF e outros usuários para

definir os requerimentos do IFTRAB2004 e elaborar os questionários que foram usados na recolha de dados. Um questionário separado foi elaborado para o trabalho infantil.

3.3.2. Dimensão da Amostra do IFTRAB2004

Os domínios geográficos de análise estabelecidos para o IFTRAB2004 foram as 10 províncias de Moçambique mais a Cidade de Maputo, como também a área urbana e rural a nível nacional. A amostra do IFTRAB2004 foi desenhada para obter resultados fiáveis para cada domínio. O IFTRAB2004 recolheu dados representativos cobrindo os quatro trimestres do ano. Assim, esta metodologia foi usada para calcular os erros padrão aproximados para as estimativas da taxa de desemprego para cada província e ao nível nacional, urbano e rural por trimestre.

A dimensão da amostra do IFTRAB2004 é apresentada no Quadro 3.3. Para obter estimativas conservadoras dos erros padrão, foi assumida uma taxa de resposta mínima de 90 por cento para agregados familiares. O Quadro 3.3 apresenta também a estimativa da média de pessoas de 15 anos e mais economicamente ativas esperadas na amostra, baseado nos dados do IAF2002/03 por domínio geográfico.

Província	Agregados familiares			Áreas de Enumeração			Pessoas de 15+ anos esperadas na amostra
	Total	Urbano	Rural	Total	Urbano	Rural	
Niassa	1.400	600	800	80	40	40	2.951
Cabo Delgado	1.720	600	1.120	96	40	56	3.445
Nampula	1.660	780	880	96	52	44	3.363
Zambezia	1.720	600	1.120	96	40	56	3.590
Tete	1.720	600	1.120	96	40	56	3.516
Manica	1.400	600	800	80	40	40	3.323
Sofala	1.640	1.080	560	100	72	28	3.785
Inhambane	1.560	600	960	88	40	48	3.299
Gaza	1.560	600	960	88	40	48	3.784
Maputo Província	1.680	1.200	480	104	80	24	3.696
Maputo Cidade	1.740	1.740	0	116	116	0	4.135
Moçambique	17.800	9.000	8.800	1.040	600	440	38.887

Quadro 3. 3 - Dimensão e repartição da Amostra do IFTRAB2004. Fonte: INE-M, IFTRAB2004

3.3.3. Listagem de agregados familiares do IFTRAB2004

Para obter uma amostra atualizada, a seleção de AF's na última etapa de amostragem foi baseada numa listagem recente dentro de cada AE da amostra. Para evitar uma seleção viciada feita por inquiridores, para o IFTRAB2004 a listagem de agregados em cada AE selecionada foi conduzida um mês antes da recolha de dados, e a seleção foi feita no escritório provincial. A tarefa de listagem para o IFTRAB2004 contou com outro pessoal distinto do de recolha de dados em cada província.

Dado que a listagem foi conduzida por uma brigada de quatro AL antes da recolha de dados para o inquérito na AE selecionada, o supervisor foi responsável de dividir o trabalho de entrevistas domiciliárias entre os inquiridores usando o mapa topográfico da área amostral.

3.3.4. Questionários do IFTRAB2004

No IFTRAB2004 foram usados 3 tipos de questionários: (i) Questionário do Agregado Familiar, (ii) Questionário Principal, para pessoas de 7 anos ou mais e (iii) Questionário Infantil.

O AF selecionado foi visitado e entrevistado. Esta entrevista foi feita ao respetivo chefe ou a um outro membro do AF que podia responder pelo chefe e ajudava a identificar todos os residentes habituais do AF (incluindo os hóspedes que residissem no AF há 6 ou mais meses).

No Questionário Principal foi recolhida informação para cada membro do AF dos 7 ou mais anos de idade. Neste questionário, foram recolhidas informações sobre atividades económicas realizadas nos últimos 7 dias anteriores à entrevista e respetivas receitas/rendimentos.

Depois de preencher este questionário, o inquiridor passava para o Questionário Infantil caso existissem no agregado pessoas dos 7 aos 17 anos de idade que realizaram alguma atividade económica nos últimos 7 dias anteriores à entrevista. Nesse Questionário Infantil foram recolhidas informações sobre percepção dos pais/encarregados, ocupação das crianças, rendimentos e horas de trabalho, saúde e segurança das crianças.

3.3.5. Organização do trabalho de campo do IFTRAB2004

Para o trabalho de campo do IFTRAB2004 e de acordo com o plano amostral, em cada AE urbana foram entrevistados 15 AF's e 20 na área rural. Cada província contou com uma brigada de recolha de dados com a seguinte composição: um supervisor, um controlador, quatro inquiridores e um motorista.

Todos os membros das brigadas beneficiaram-se de uma formação conduzida pela Direção Central do IFTRAB2004 que compreendeu sessões teóricas e práticas sobre a condução da entrevista, entrevistas simuladas na sala e ensaio no terreno.

3.3.6. Procedimentos usados na Condução de entrevistas do IFTRAB2004

De igual modo como se descreveram os procedimentos usados na condução de entrevistas do IDS2003, no âmbito do IFTRAB2004 foram usados os mesmos procedimentos. Durante a formação de pessoal de campo salientaram-se vários aspetos dentre os quais a necessidade de transformar a entrevista em conversa amigável entre duas partes e, ainda, que esta pudesse ser conduzida como uma conversa normal, o que implica a observância de regras básicas para o seu êxito.

Os mesmos aspetos mais importantes que se tiveram em conta durante as entrevistas do IDS2003 foram reforçados no âmbito do IFTRAB2004, nomeadamente, o acesso ao entrevistado, a confidencialidade das respostas, a neutralidade a observar durante a entrevista, o controlo que se deve ter da entrevista, como proceder ao se lidar com pessoas indecisas, a língua usada durante a entrevista e o fecho da entrevista.

3.3.7. Procedimento para Contactos aos agregados familiares

Identicamente aos procedimentos para contactos usados no IDS2003, o IFTRAB2004 também recoendou os mesmos procedimentos atendendo que se trata de inquérito demográfico. Ainda, foi salientado que o IFTRAB2004 teria êxito se cada membro da brigada compreendesse e seguisse corretamente os procedimentos metodológicos de campo.

Ao pessoal de campo era-lhe exigido que se familiarizassem com os procedimentos, e que soubessem solucionar possíveis problemas derivados da complexidade do próprio trabalho de campo.

Uma das primeiras tarefas da brigada nas AE's seleccionadas foi o contacto direto com as autoridades administrativas locais (os secretários, chefes das aldeias, dos bairros, dos quarteirões). Portanto, antes de se começar a entrevista, devia ter-se certeza de que os líderes locais da AE estavam informados e sabiam dos trabalhos do inquérito. Por outro lado, o controlador ou supervisor teria trabalhado nesse sentido de mobilizar à participação da comunidade abrangida.

Nas áreas urbanas, a carga de trabalho do IFTRAB2004 foi de 15 AF's o que significou que cada inquiridor visitava e entrevistava 3 a 4 AF's. Nas áreas rurais foi de

20 AF's e cada inquiridor entrevistava 5. O tempo de permanência total da brigada em cada área de enumeração foi de 2 dias.

3.4. PROCEDIMENTOS METODOLÓGICOS DO IOF2009

Os procedimentos metodológicos que a seguir se apresentam são parte integrante de três documentos: Manual de Inquiridor do Inquérito ao Orçamento Familiar de 2009 elaborado por INE-M (2008), *Sampling Recommendations for the Mozambique* Inquérito ao Orçamento Familiar de 2008/09 produzido por Megill (2008) e ainda, Recomendação para o desenho da amostra do Inquérito aos Agregados Familiares de 2002/03 produzido também por Megill (2002).

O INE-M realizou o IAF em 2002/03. O principal objetivo do IAF2002 foi de obter dados nacionais e provinciais sobre as despesas familiares, rendimentos, características demográficas e socioeconómicas. Esses dados foram utilizados para estudos relacionados com a pobreza principalmente o Plano de Ação para a Redução da Pobreza Absoluta (PARPA), atualizar o cabaz, atualizar os pesos para o Índice de Preços ao Consumidor e ainda para as contas nacionais (Megill, 2002).

Decorridos cerca de cinco anos desde 2003, houve necessidade de se atualizar os dados do IAF2002 a partir de outro inquérito aos agregados familiares. Nisto, o INE-M realizou o IOF2009 durante 12 meses de Julho de 2008 a Junho de 2009.

3.4.1. Objetivos do IOF2009

Identicamente ao IAF2002, o principal objetivo do IOF2009 foi de obter dados nacionais e provinciais representativos sobre as despesas familiares, rendimentos e características socioeconómicas dos AF's. Por outro lado, os dados do IOF2009 foram

utilizados para atualizar o cabaz e os pesos para o Índice de Preços ao Consumidor e, ainda para as contas nacionais de Moçambique.

3.4.2. A Amostragem do IOF2009

O universo definido para o IOF2009 foi composto por todos AF's residentes em habitações particulares de todo o país. Excluíram-se os AF's em habitações coletivas, os 'Sem casa' e diplomatas residentes nas embaixadas/representações que juntos representam menos de um por cento do total da população moçambicana (INE-M, Censo 2007).

A amostra do IOF 2009 foi extraída da base de sondagem recente obtida a partir dos dados de atualização cartográfica e dos dados do Recenseamento Geral de População e Habitação de 2007. Esses dados incluem o número de habitações, número de AF's, número de área de enumeração, todos repartidos por províncias e distritos.

3.4.3. Dimensão da amostra do IOF2009

A amostra do IOF2009 abrangeu 10.848 AF's em todo o país, dos quais 5.232 são das zonas urbanas e 5.616 são das rurais. O número de agregados familiares que foram entrevistados em cada província varia entre 768 e 1.584 (Quadro 3.4). A dimensão final da amostra foi ajustada à taxa de resposta de 90% para AF's em cada uma das províncias. A amostra permitiu obter estimativas fiáveis para os domínios: Nacional, zona de residência (Urbano, Rural), Regional (Norte, Centro e Sul) e Provincial.

Ainda de acordo com o Quadro 3.4, os 10.848 AF's foram selecionados em 1.060 áreas de enumeração distribuídas de forma proporcional por cada província e zona de residência urbana e rural.

Província	Agregados familiares			Áreas de Enumeração		
	Total	Urbano	Rural	Total	Urbano	Rural
Niassa	816	384	432	80	32	48
Cabo Delgado	780	240	540	80	20	60
Nampula	1.584	576	1.008	160	48	112
Zambézia	1.524	336	1.188	160	28	132
Tete	768	192	576	80	16	64
Manica	804	336	468	80	28	52
Sofala	852	528	324	80	44	36
Inhambane	804	336	468	80	28	52
Gaza	816	384	432	80	32	48
Maputo Província	900	720	180	80	60	20
Maputo Cidade	1.200	1.200	0	100	100	0
Moçambique	10.848	5.232	5.616	1.060	436	624

Quadro 3. 4 – Dimensão e repartição da amostra do IOF2009. Fonte: INE-M, IOF2009

3.4.4. Listagem de agregados familiares do IOF2009

Novas listas de AF's foram elaboradas em todas as AE's amostrais selecionadas. Para apoio do processo de listagem foram utilizados mapas topográficos das AE's e para localização das habitações dos AF's, utilizando GPS²¹, foram recolhidas as respetivas coordenadas geográficas.

3.4.5. Questionários do IOF2009

O IOF2009 utilizou três tipos de questionários:

- (i) Questionário do Agregado familiar. Neste questionário foram recolhidas informações básicas de cada pessoa, o sexo, a idade, o nível de instrução, estado civil, sua atividade económica. Ainda neste

²¹ O Sistema de Posicionamento Global (GPS) é um espaço baseado em [navegação por satélite](http://en.wikipedia.org/wiki/Global_Positioning_System). O sistema fornece a localização e informações em todo tempo e em qualquer lugar da Terra. Fonte: http://en.wikipedia.org/wiki/Global_Positioning_System.

questionário foram recolhidas informações sobre características da habitação, tipo de saneamento, fonte de água, posse de bens duráveis.

- (ii) Questionário das Despesas Diárias do Agregado. Neste questionário foram apuradas as despesas referentes aos produtos alimentares; bebidas alcoólicas e não alcoólicas; tabaco e narcóticos; refeições e refrigerantes consumidos em restaurantes, cafés, bares, cantinas; gastos em transporte local. Ainda este questionário incluiu o valor do autoconsumo e as receitas em espécie.
- (iii) Questionário das Despesas e Receitas. Este questionário serviu para recolha de dados sobre despesas mensais, despesas anuais, receita, transferências e posse de bens duráveis.

3.4.6. Organização do trabalho de campo de IOF2009

O trabalho de campo do IOF2009 foi realizado de acordo com o respetivo cronograma, e cada membro da brigada devia compreender e seguir corretamente os procedimentos de campo. Em particular, aos inquiridores era-lhes exigido que estivessem familiarizados com os procedimentos de campo, para poderem solucionar corretamente os problemas que podiam surgir durante o trabalho de campo. A recolha de dados do IOF2009 foi feita durante 12 meses, divididos em 26 períodos de 10 dias cada. Em cada período, o inquiridor fazia a recolha de dados em 12 AF's nas zonas urbanas e em 8 nas zonas rurais. Cada AF selecionado foi entrevistado. Essa entrevista foi feita ao chefe do AF ou a outro membro do agregado que respondia pelo chefe e ajudava a identificar todos os residentes habituais do AF. Os visitantes cujo período de permanência fosse de 6 meses ou mais no AF foram inclusos na amostra do IOF2009.

3.4.7. Procedimentos aplicados na Condução de entrevistas do IOF2009

De igual modo, os mesmos procedimentos usados na condução de entrevistas do IDS2003 e do IFTRAB2004 foram usados também no IOF2009.

Os aspetos mais importantes que se tiveram em conta durante as entrevistas do IDS2003 e do IFTRAB2004 são os mesmos que foram reforçados no IOF2009, nomeadamente, o acesso ao entrevistado, a confidencialidade das respostas, a neutralidade a observar durante a entrevista, o controlo que se deve ter da entrevista, como proceder ao se lidar com pessoas indecisas, a língua usada durante a entrevista e o fecho da entrevista.

3.4.8. Procedimento para Contactos aos agregados familiares

Identicamente aos procedimentos para contactos aos AF's usados no IDS2003 e no IFTRAB2004, a metodologia do IOF2009 aplicou os mesmos procedimentos, atendendo que também trata-se de inquérito demográfico.

3.5. TÉCNICAS E MÉTODOS DE ANÁLISE DOS DADOS

Esta secção visa apresentar sumariamente as técnicas, os métodos e os procedimentos a serem utilizados na análise de dados para o alcance dos objetivos desta dissertação. Segundo Reis (2010), a análise de dados é a tentativa de evidenciar as relações existentes entre o fenómeno estudado e outros fatores. Portanto, o objetivo da análise e tratamento de dados consiste em verificar se as informações obtidas validam as hipóteses formuladas, isto é, se os resultados observados correspondem aos resultados esperados.

3.5.1. Semelhanças nos Planos Amostrais dos três inquéritos

Nas Secções 3.2 a 3.4 foram descritos os procedimentos metodológicos do desenho das amostras dos três inquéritos ora anunciados. Com esta tarefa pretendeu-se avaliar as (dis)semelhanças existentes. A descrição dos inquéritos ora apresentada remete-nos a concluir que os desenhos amostrais são semelhantes, os três inquéritos utilizam o mesmo tipo de amostragem probabilística, a amostra é selecionada em três etapas: a primeira centrada na seleção de UPA's com probabilidade proporcional a dimensão da UPA em termos de AF's, a segunda que consiste na seleção de AE's com probabilidades iguais dentro das UPA's e, a terceira centrada na seleção aleatória de AF's nas AE's amostrais com probabilidades iguais. Outra semelhança está ligada ao processo de listagem dos AF's nas AE's, pois os agregados selecionados na terceira etapa são de entre os listados momentos antes da recolha de dados. Outra semelhança ainda, prende-se com a estratificação da amostra. Ou seja, os estratos definidos para os três inquéritos são os mesmos: estrato urbano e rural em cada província. Em cada estrato foram definidos substratos. No estrato urbano foram classificadas as zonas urbanas em cidades capitais, cidades com mais de 20.000 AF's e resto urbano²². Ainda outros substratos foram definidos para o estrato rural: as zonas agro-ecológicas²³. As amostras de UPA's, de AE's e de AF's foram selecionadas de forma independente para cada substrato. A amostra global foi a soma das sub amostras correspondentes aos substratos de cada província.

A dissemelhança reside fundamentalmente nos objetivos de cada inquérito – os já descritos nas Subsecções 3.2.1, 3.3.1 e 3.4.1 para o IDS2003, IFTRAB2004 e IOF2009, respetivamente.

²² De acordo com o Censo 1997, o 'Resto urbano' é constituído por cidades médias (menos de 20.000 AF's), cidades pequenas e vilas. Para a Cidade de Maputo, capitais provinciais e cidades grandes foram definidos ainda outros substratos socioeconómicos com base nas características da habitação e nível de educação dos membros do agregado familiar.

²³ No Censo 1997, as Zonas Agro-ecológicas foram definidas como sendo espaços territoriais com condições climáticas para a prática da agricultura em sequeiro que predominantemente são os planaltos, as zonas de altitude intermédia, as zonas costeiras de altitude baixa, as zonas secas e muito secas.

Assim, atendendo à grande semelhança da metodologia aplicada para os três inquéritos, no que diz respeito ao desenho amostral, à condução das entrevistas, aos procedimentos de campo, assume-se que as diferenças que poderão ser observadas nos resultados da análise das BD's dos inquéritos serão devidas à aleatoriedade do fenómeno em estudo ou à incidência dos fatores da não-resposta nos AF's e, para isso, com recurso à inferência estatística, vão ser elaborados testes de significância das taxas de respostas, das proporções ou contagens associadas aos fatores da não-resposta referentes à cada inquérito com vista a permitir comparações estatisticamente válidas.

3.5.2. Análise das BD's e os Testes de Significância

Conforme foi referenciado no último parágrafo da Secção 3.1, os dados dos três inquéritos constam das respetivas BD's. Uma análise de cada uma BD's vai ser feita de modo a determinar as variáveis que estão ligadas ao controlo dos erros de não-resposta. Com base nessas variáveis e usando as Equações 2.8 a 2.13 vão ser calculadas taxas de resposta de cada um dos inquéritos. A tabela de frequências das variáveis previamente determinadas e identificadas na BD fornecerá as quantidades para o numerador e o denominador da Equação 2.12. A escolha desta equação foi com base na taxa de resposta que se pretende calcular, pois esta equação é usada em situações em que todos os casos são de elegibilidade conhecida face aos objetivos do inquérito. No caso dos três inquéritos ora descritos, todos os AF's residentes em habitações particulares são elegíveis e equiprováveis.

Teste Binomial

Com recurso a análise inferencial, incorporado no *software* estatístico *PASW Statistics* versão 18, vão ser realizados testes de significância para avaliar o quão as taxas de resposta de cada um dos três inquéritos, estatisticamente, correspondem à taxa de resposta na população de AF's cuja taxa de resposta mínima é 90 por cento – este parâmetro foi assumido no plano amostral de cada um dos três inquéritos. Assim, o teste binomial vai ser realizado, pois segundo Marôco (2010), este teste é geralmente utilizado para testar a ocorrência de uma das duas realizações de uma variável dicotómica, i.é, para testar a proporção dessa ocorrência no total de

ocorrências registadas. Atendendo que a dimensão da amostra nos três inquéritos é grande [se $n > 20$ e $np > 7$, segundo Marôco (2010)], a distribuição binomial pode ser aproximada à normal. Portanto, em alternativa ao uso da distribuição de probabilidades da distribuição binomial, pode-se usar a distribuição de probabilidades da normal.

Assim, para efeitos do teste binomial e recorrendo a variável ‘Resultados da entrevista’, em cada BD vai ser criada outra variável binária cujos valores serão “1” para os “respondentes, com entrevistas completas” e “2” para os “não respondentes, os casos classificados como não-resposta”. A proporção do teste binomial vai ser $p=0,90$ nos três inquéritos, pois pretende-se testar se as amostras foram extraídas de população onde a taxa de resposta mínima é 90 por cento. Os resultados do teste binomial vão apoiar na decisão se as taxas de respostas são estatisticamente diferentes de 90 por cento ou não. Caso não sejam estatisticamente superiores a 90 por cento – rejeição da hipótese nula – concluir-se-á que os erros de não-resposta produzem uma taxa de resposta inferior a 90 por cento.

Teste de Aderência do Qui-quadrado

Por um lado, vai ser preciso avaliar a aderência da distribuição de frequências associadas a cada uma das categorias da variável nominal ‘Resultado da entrevista’ nos três inquéritos através de um teste de significância de Qui-quadrado. De acordo com Guimarães e Cabral (2011), o teste do Qui-quadrado para qualidade do ajuste permite também avaliar a aderência entre uma distribuição de frequências associada a uma ou duas amostras independentes constituídas por observações expressas numa qualquer escala e uma distribuição teórica. Ainda de acordo com estes autores, os requisitos exigidos para a realização deste teste são os de que as amostras sejam aleatórias e tenham uma dimensão mínima adequada²⁴.

²⁴ Tanto para o teste de ajuste quanto para o de independência, a dimensão da amostra deve ser tal que as frequências esperadas no teste Qui-quadrado não sejam inferiores a 5 (Guimarães e Cabral, 2011:295). Sobre este assunto, Marôco (2010) recomenda que o teste Qui-quadrado só pode ser aplicado com rigor quando sejam verificadas as seguintes condições: dimensão da amostra seja superior a 20 casos; todas as frequências esperadas sejam superiores a 1; pelo menos 80% das frequências esperadas sejam superiores ou iguais a 5.

Assim, o teste Qui-quadrado permite testar se a distribuição dos elementos da amostra global pelas classes ou categorias de uma determinada variável é aleatória ou não. Se a distribuição dos elementos observados for aleatória, isso é equivalente a afirmar que as distribuições nas classes das variáveis são independentes das amostras ou grupos (ou também das classes de uma segunda variável). No primeiro cenário, o teste de Qui-quadrado é designado de teste de homogeneidade ou aderência; enquanto que no segundo cenário designa-se por teste de independência. No teste de independência, os dados (contagens) organizam-se, geralmente, em tabelas de contingência (Marôco, 2010).

Para os propósitos desta dissertação, vão ser considerados os três inquéritos separadamente envolvendo a variável do controlo da não-resposta previamente determinada. A estatística do teste, neste caso, é a medida global do ajuste entre as contagens observadas nas amostras e as respetivas contagens esperadas (Guimarães e Cabral, 2011) sob hipótese nula de que *“Na população de onde se extraíu a amostra, não existem diferenças entre as contagens que se distribuem pelas categorias da variável nominal ‘Resultado da entrevista’ previamente determinada”*. Ou seja, o resultado deste teste de aderência vai apoiar na decisão sobre se, na população de onde é extraída a amostra²⁵, todas as categorias têm a mesma proporção de observações. Portanto, com este teste poder-se-á decidir ou não sobre fatores que influenciam a taxa de resposta por inquérito e, conseqüentemente, concluir-se que os erros de não-resposta são devidos a tais fatores.

Os fatores a serem identificados vão ser escalonados de acordo com a magnitude das suas frequências com intuito de identificar os principais. Esta técnica estatística vai ser aplicada para cada um dos três inquéritos.

Teste de Independência do Qui-quadrado

Por outro lado, tem de se avaliar se os fatores de não-resposta identificados são ou não independentes do tipo de inquérito, através de mais um teste de significância de Qui-quadrado. Segundo Marôco (2010), o teste de independência do Qui-quadrado

²⁵ Este teste de aderência do Qui-quadrado compara as frequências observadas e esperadas em cada categoria testando se cada categoria contém a mesma proporção ou outra especificada pelo pesquisador. Em geral a hipótese nula é formulada nos seguintes termos: *‘As frequências registadas nas categorias da variável não diferem consoante as categorias’* contra a alternativa *‘Pelo menos uma das categorias tem frequência diferente das demais’* (Guimarães e Cabral, 2011).

serve para testar se duas ou mais populações (ou grupos) independentes diferem relativamente a uma determinada característica, isto é, se a frequência com que os elementos da amostra se repartem pelas categorias de uma variável nominal é ou não aleatória.

Para os propósitos desta dissertação, na BD conjunta, vão ser cruzadas duas variáveis ‘Tipo de inquérito’ e ‘Resultado da entrevista’, numa tabela de contingência, sob hipótese nula de que *“Nas populações de onde foram extraídas as amostras a distribuição das contagens pelas categorias é independente da variável”*. O resultado deste teste de independência vai apoiar na decisão se os fatores dos erros da não-resposta são independentes ou não do tipo de inquérito. Ou dito de outra forma, se as contagens observadas nas categorias da variável ‘Resultado da entrevista’ são devidas à aleatoriedade ou dependem do ‘Tipo de inquérito’ conduzido pelo INE-M na Cidade de Maputo.

A identificação ou não dos principais fatores vai permitir que sejam tomadas medidas que visam prevenir ou controlar a ocorrência de erros não amostrais ligados à não-resposta. Adicionalmente, caso se decida pela não independência entre o ‘Resultado da entrevista’ e o ‘Tipo de inquérito’ poder-se-á recomendar medidas específicas que possam ser tomadas em conta para prevenir ou controlar a ocorrência de erros de não-resposta.

Nesta dissertação, todos os testes de hipóteses vão ser efetuados sob um nível de significância de 5% e os resultados das análises estatísticas feitas vão ser apresentados em forma de quadros, gráficos ou diagramas comparativos entre os três inquéritos para a Cidade de Maputo.

4. ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

Neste capítulo é feita a descrição de BD's dos três inquéritos (IDS2003, IFTRAB2004, IOF2009) e, mais adiante vai ser feita a apresentação e discussão de resultados.

Conforme foi referenciado, esta dissertação é elaborada com base nos dados de inquéritos nacionais para a Cidade de Maputo. O Quadro 4.1 mostra a distribuição da dimensão das amostras de AF's e das AE's planificadas nos três inquéritos.

Inquérito analisado	Agregados Familiares	Áreas de Enumeração
IDS2003	1.248	52
IFTRAB2004	1.740	116
IOF2009	1.200	100

Quadro 4. 1 – Dimensão das amostras dos três inquéritos na Cidade de Maputo. Fonte: IDS2003, IFTRAB2004, IOF2009

De acordo com os respetivos desenhos amostrais, o trabalho de recolha de dados tanto o IFTRAB2004 quanto ao IOF2009 decorreu durante um período de 12 meses enquanto que o IDS2003 foi em 4 meses.

4.1. DESCRIÇÃO DAS BASES DE DADOS USADAS

Conforme foi exposto no último parágrafo da Secção 3.1, para o alcance dos objetivos desta dissertação, vão ser analisados dados recolhidos em cada um dos três inquéritos. Esses dados constam das respetivas bases. Uma breve descrição do desenho das BD's é apresentada. Nessas BD's, os casos são os agregados familiares. Em cada agregado foi entrevistado, face-a-face, o respetivo chefe ou, na sua ausência, foi entrevistado um adulto de 18 ou mais anos de idade com capacidade de declarar sobre o agregado.

4.1.1. Base de Dados do IDS2003

A BD global de agregados familiares do IDS2003 é constituída por um total de 102 variáveis. Para esta dissertação, foram consideradas as variáveis cuja análise interessa e seus conteúdos estão ligados aos objetivos preconizados. A Figura 4.1 apresenta uma parte dessas variáveis com destaque nas ligadas ao resultado da entrevista junto ao agregado familiar.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	QIDSNUM	Numeric	3	0	Numero do conglomer...	None	None	7	Right	Scale
2	QAFNUM	Numeric	3	0	Numero do agregado	None	None	6	Right	Scale
3	Agreg_fam	Numeric	10	0	Agregado familiar Con...	None	None	11	Right	Scale
4	Agregado	Numeric	3	0	Filtrar Agregado = 1	None	None	8	Right	Scale
5	QPROVIN	Numeric	2	0	Provincia	{1, Niassa}...	None	7	Right	Nominal
6	QTYPE	Numeric	1	0	Urbano - Rural	{1, Urbano}...	None	5	Right	Nominal
7	QLOCAL	Numeric	1	0	Localidade	{1, Ciudad g...	None	6	Right	Scale
8	QQTYPE	Numeric	1	0	Tipo do inquerito	{1, Home}...	None	6	Right	Nominal
9	QINT	Numeric	8	0	Data	None	None	8	Right	Scale
10	QINTD	Numeric	2	0	Dia	None	None	5	Right	Scale
11	QINTM	Numeric	2	0	Mes	None	None	5	Right	Scale
12	QINTY	Numeric	4	0	Ano	None	None	5	Right	Scale
13	QINTNUM	Numeric	3	0	Numero Inquirido(a)	None	None	7	Right	Scale
14	ENTRCOMPLET	Numeric	2	0	Entrevista completa	{0, Não}...	None	5	Right	Nominal
15	QRESULT	Numeric	2	0	Resultado	{1, Complet...	None	7	Right	Nominal
16	QVISITS	Numeric	1	0	Número total de visitas	None	None	7	Right	Scale
17	QSUPERV	Numeric	3	0	Supervisor	None	None	7	Right	Scale
18	QFEDIT	Numeric	3	0	Controlador	None	None	6	Right	Scale
19	QOEDIT	Numeric	2	0	Revisto No gabinete por	None	None	6	Right	Scale
20	QKEYER	Numeric	2	0	Digitado por	{1, MERICI...	None	6	Right	Scale
21	QKEYER2	Numeric	2	0	Redigitado por	{1, MERICI...	None	7	Right	Scale
22	QINTC	Numeric	4	0	Date of interview (CMC)	None	None	5	Right	Scale

Figura 4. 1 – Uma parte das variáveis da BD de AF's do IDS2003, em SPSS.

Nessa BD, a variável ligada ao controlo do decurso da recolha de dados e dos erros de não-resposta é identificada por QRESULT que traduz os possíveis resultados obtidos ao nível do AF incluído na amostra do IDS2003. Esta variável cujas categorias são *Completo*, *Incompleto*, *Recusa Total*, *Recusa Durante a Entrevista*, *Todo o Agregado Ausente*, *Agregado Incapacitado* e *Outro* foi também incluída nos questionários dos AF's dos restantes inquiridos.

Para esta variável foram obtidas as frequências apresentadas no Quadro 4.2. De acordo com o plano de amostragem do IDS2003, foram planificados um total de 1.248 agregados (ver Quadro 4.1). Em contrapartida, foram visitados 941 AF's devido à limitações de ordem financeira.

Resultado no Agregado	Frequência	Perce tagem	Percentagem acumulada
Completo	730	77,6	77,6
Ausente	171	18,2	95,7
Recusa Total	18	1,9	97,7
Incompleta	3	0,3	98,0
Incapacitada	14	1,5	99,5
Outro	5	0,5	100,0
Total	941	100,0	-

Quadro 4. 2 – Resultados das visitas ao AF para entrevistas do IDS2003. Fonte: Elaboração própria

Dos 941 agregados que foram visitados, em 730 obteve-se entrevistas Completas (77,6%), os restantes 211 distribuem-se entre os Ausentes (18,2%), Recusa total (1,9%), Incapacitados (1,5%), Entrevista Incompleta (0,3%) ou Outras razões (0,5%).

Com vista a aumentar a colaboração dos agregados foram efetuadas visitas de insistência principalmente nos agregados ausentes. Estas tentativas resultaram que uns fossem entrevistados na 1ª visita, outros na 2ª e, ainda outros na 3ª visita. O Quadro 4.3 mostra o número de visitas efetuadas e os respetivos agregados abordados.

Número total de visitas	Frequência	Perce tagem	Percentagem acumulada
1ª visita	707	75,1	75,1
2ª visita	96	10,2	85,3
3ª visita	138	14,7	100,0
Total	941	100,0	-

Quadro 4. 3 – Frequências de AF's segundo as visitas efetuadas durante o IDS2003. Fonte: Elaboração própria

Usando as notações introduzidas na Subsecção 2.6.1 referentes às taxas de resposta definidas pela AAPOR (2011), pode-se calcular a taxa de resposta do IDS2003 usando a Equação 2.12.

O uso desta equação prende-se com o fato de que todos os agregados familiares são de elegibilidade conhecida (AAPOR, 2011). Adicionalmente, as entrevistas

incompletas são consideradas nesta dissertação como não-respostas, pois segundo Biemer e Lyberg (2003:80):

There are no clear guidelines to say when a unit has answered enough of the items on a survey to be considered as a responding unit with extensive item nonresponse, as opposed to a unit nonrespondent. (...) Usually, criteria are developed to determine how many or which key items on the questionnaire need to be completed before a questionnaire can be classified as completed. Then a questionnaire that does not meet these criteria is called incomplete and the unit is classified as a nonrespondent.

Assim, através da Equação 2.12 tem-se:

$$RR_{IDS2003} = \frac{730}{(730 + 3) + (171 + 18 + 14 + 5)} = \frac{730}{941} = 0,776$$

O Quadro 4.2 mostra este mesmo resultado da taxa de resposta. Portanto, no IDS2003, a taxa de resposta, sem considerar as entrevistas parciais no numerador, é de 77,6%. Esta taxa é influenciada pelos erros de não-resposta. Quanto maior for o número de casos de não-resposta menor é a taxa de resposta.

Com base no plano do desenho da amostra do IDS2003, a dimensão da amostra foi calculada com pressuposto de que no universo de onde se extraíu a amostra, a taxa de resposta dos AF's na Cidade de Maputo é pelo menos de 90%. Pode-se aplicar um teste binomial usando-se a distribuição de probabilidades da normal, conforme recomenda Marôco (2010), pois a dimensão da amostra é 941 casos.

Na BD do IDS2003 foi criada uma variável dicotómica ENTRCOMPLET para classificar se no agregado obteve-se entrevista completa ou não e cujas categorias são 1=Sim e 0=Não. Com esta nova variável as hipóteses estatísticas por testar são formuladas nos seguintes termos:

H_0 : A taxa de resposta na população de AF's da Cidade de Maputo é pelo menos 90 por cento ($p \geq 0,90$);

H_1 : A taxa de resposta na população de AF's da Cidade de Maputo é inferior a 90 por cento ($p < 0,90$).

Portanto, para avaliar a significância estatística da taxa de resposta de AF's no IDS2003 na Cidade de Maputo, recorreu-se ao teste binomial implementado no *software* de análise de dados PASW Statistics (v.18; IBM SPSS, Chicago, IL) como descrito em Marôco (2010). Considerou-se $\alpha=0,05$.

Resultados

Na presente dissertação, da análise dos dados do IDS2003 feita, a taxa de resposta de agregados familiares foi de 77,6% contra os 22,4% considerados não-resposta – esta distribuição é apresentada no Gráfico 4.1. Contudo, a análise estatística inferencial indica que a taxa de resposta é significativamente diferente de 90 por cento, concretamente a taxa é inferior a 90 por cento (*valor* $p=0,000 < \alpha=0,05$; $n=941$). Os resultados do teste de significância podem ser vistos no Anexo A5.

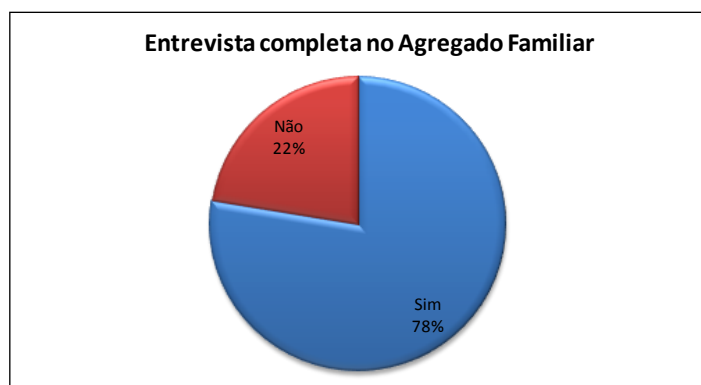


Gráfico 4. 1 – Distribuição percentual da variável ENTRCOMPLET do IDS2003.

Fonte: Elaboração própria

Deste modo, os resultados do teste realizado remetem à rejeição da hipótese nula (H_0) ao nível de significância de 5%. Portanto, a amostra do IDS2003 foi extraída de uma população cuja taxa de resposta não é de pelo menos 90 por cento. Ou seja, na população aonde foi extraída a amostra de AF do IDS2003, a taxa de resposta é inferior a 90 por cento.

Fatores da Não-Resposta no IDS2003

Pelo Quadro 4.2, a amostra do IDS2003 pode ser repartida em duas sub amostras: os respondentes ($n_1=730$) e os não respondentes ($n_2=211$). Uma análise particular é feita à volta dos não respondentes. O Quadro 4.4 mostra que 81% dos

agregados não respondentes fizeram-no por Ausência de todo agregado no período das entrevistas, 8,5% recusaram-se totalmente em prestar entrevistas, em 6,6% houve membros residentes incapacitados em prestar declarações. Houve entrevistas incompletas (algumas perguntas do questionário não respondidas) e os restantes 2,4% são agregados que não responderam por outros motivos (por exemplo, infelicidade no AF, língua usada).

Fatores de Não-Resposta do AF	Frequência	Porcentagem	Porcentagem acumulada
Ausência do AF	171	81,0	81,0
Recusa Total	18	8,5	89,6
Entrevista Incompleta	3	1,4	91,0
Incapacidade	14	6,6	97,6
Outro	5	2,4	100,0
Total	211	100,0	-

Quadro 4. 4 – Fatores de Não-Resposta no IDS2003

Com relação às frequências apresentadas no Quadro 4.4, pode-se realizar um teste para avaliar a significância das diferenças entre as frequências. Nisto, o Quadro 4.5 mostra a distribuição de frequências observadas e esperadas do teste Qui-quadrado. Este teste, também chamado teste de homogeneidade, compara as frequências observadas e as esperadas em cada categoria para testar se todas as categorias contêm a mesma proporção ou frequência de valores ou não.

Fatores de Não-Resposta do AF	Frequência Observada	Frequência Esperada	Resíduo
Ausência do AF	171	42,2	128,8
Recusa Total	18	42,2	-24,2
Entrevista Incompleta	3	42,2	-39,2
Incapacidade	14	42,2	-28,2
Outro	5	42,2	-37,2
Total	211	-	-

Quadro 4. 5 – Distribuição de frequências esperadas dos não respondentes no IDS2003

Deste modo, as hipóteses do teste podem ser formuladas nos seguintes termos:

H_0 : Na população dos AF's não respondentes as frequências observadas não diferem das esperadas nas respetivas categorias.

H_1 : Na população dos AF's não respondentes as frequências observadas diferem das esperadas nas respectivas categorias.

Portanto, sob hipótese nula todos os fatores são homogêneos e apresentam igual distribuição de frequências na população dos não respondentes.

Resultados do teste

A análise inferencial feita sobre as frequências do Quadro 4.5, indica que as frequências observadas diferem das esperadas ($\chi^2=495,043$; $gl=4$; valor $p=0,000 < \alpha=0,05$). Os resultados deste teste de significância constam do Anexo A6. Portanto, os fatores da não-resposta no IDS2003 diferem entre si. Em termos de ordenação pode-se inferir com 95% de confiança que o principal fator é a Ausência dos AF's das suas habitações no período das entrevistas (81,0%), em seguida é a Recusa total (8,5%), em terceiro é a Incapacidade dos membros do AF (6,6%) e, finalmente, Outros fatores (2,4%).

4.1.2. Base de Dados do IFTRAB2004

A BD de AF's do IFTRAB2004 é constituída por um total de 457 variáveis. Para os objetivos deste trabalho, foram consideradas as variáveis cujos conteúdos estão ligados ao resultado da entrevista, nomeadamente: RESULTADO e VISITAS. Além destas, foi criada outra variável binária (ENTRCOMPLET) idêntica a que se criou na BD do IDS2003. A Figura 4.2 apresenta uma parte das 457 variáveis.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	AE	Numeric	4	0	Area de Enume...	None	None	4	Right	Scale
2	AF	Numeric	3	0	Numero do Agr...	None	None	3	Right	Scale
3	Agregado_1	Numeric	2	0	Somente AF=1	None	None	8	Right	Scale
4	Região	Numeric	8	2	Região	{1,00, Norte...	None	8	Right	Nominal
5	PROV	Numeric	2	0	Provincia	{1, Niassa}...	None	6	Right	Nominal
6	DISTR	Numeric	2	0	Distrito	None	None	5	Right	Scale
7	POSTO	Numeric	2	0	Posto Administ...	{99, }...	None	6	Right	Scale
8	LOCAL	Numeric	2	0	Localidade	{99, }...	None	5	Right	Scale
9	URB_RUR	Numeric	1	0	Urbano/Rural	{1, Urbano}...	None	8	Right	Nominal
10	DATA	String	8	0	Data	None	None	8	Left	Nominal
11	DIA	Numeric	2	0	Dia	{98, Não sa...	99	3	Right	Scale
12	MES	Numeric	2	0	Mes	{1, Janeiro}...	99	4	Right	Scale
13	ANO	Numeric	4	0	Ano	{2004, 2004}...	None	4	Right	Scale
14	CODIGO	Numeric	3	0	Codigo	None	None	8	Right	Scale
15	RESULTADO	Numeric	2	0	Resultado	{1, Complet...	None	9	Right	Nominal
16	ENTRCOMPLET	Numeric	2	0	Entrevista com...	{0, Não}...	None	13	Right	Nominal
17	VISITAS	Numeric	1	0	Visitas	None	None	7	Right	Scale
18	SituacaoA	Numeric	8	2	Situação do Ag...	{1,00, Inquiri...	None	9	Right	Nominal
19	TMEMBROS	Numeric	2	0	TMembros	None	None	9	Right	Scale
20	TMEMBROS_R	Numeric	8	2	Total de membr...	{1,00, 1 até ...	None	9	Right	Scale
21	TCRIA	Numeric	2	0	TCria	None	None	5	Right	Scale
22	INQUIRIDO	Numeric	2	0	Inquirido	None	None	9	Right	Nominal
23	RGAB	Numeric	2	0	RGab	{1, Eugenio}...	None	5	Right	Scale
24	DIGITADOR	Numeric	2	0	Digitador	{1, Eugenio}...	None	9	Right	Scale

Figura 4. 2 – Uma parte das variáveis da BD de AF's do IFTRAB2004, em SPSS.

De acordo com o plano de amostragem do IFTRAB2004, foram planejados um total de 1.740 agregados (ver Quadro 4.1) e todos foram visitados. As visitas de insistência foram tidas em conta com vista a garantir maior cobertura da amostra. O Quadro 4.6 mostra a distribuição percentual dos agregados por categoria de visita.

Número total de visitas	Frequência	Percentagem	Percentagem acumulada
1ª visita	1587	91,2	91,2
2ª visita	83	4,8	96,0
3ª visita	68	3,9	99,9
4ª visita	2	0,1	100,0
Total	1740	100,0	-

Quadro 4. 6 – Frequências de AF's segundo as visitas efetuadas durante o IFTRAB2004. Fonte: Elaboração própria

Pelo Quadro 4.6, foram necessárias quatro visitas na tentativa de obter cooperação dos AF's durante as operações de recolha de dados do IFTRAB2004. Apesar das quatro visitas, o resultado da entrevista não foi completo nos 1740 AF's. Houve casos em que todos os membros do AF não se encontravam na habitação durante o período de recolha de dados no conglomerado – classificados como 'Todo o AF ausente' (2,6%) – e, ainda as recusas tiveram seu peso neste inquérito (6,0%). O Quadro 4.7 mostra a distribuição de frequências da variável RESULTADO.

Resultado no Agregado	Frequência	Percentagem	Percentagem acumulada
Completo	1570	90,2	90,2
Recusa total	104	6,0	99,7
Todo o AF ausente	45	2,6	93,6
Incompleto	14	0,8	91,0
Recusa durante a entrevista	2	0,1	93,7
Outro	5	0,3	100,0
Total	1740	100,0	-

Quadro 4. 7 – Resultados das visitas ao AF para entrevistas do IFTRAB2004. Fonte: Elaboração própria

De igual forma, usando a Equação 2.12, pode-se calcular a taxa de resposta do IFTRAB2004.

$$RR_{IFTRAB\ 2004} = \frac{1570}{(1570 + 14) + (104 + 45 + 2 + 5)} = \frac{1570}{1740} = 0,902$$

Este resultado é apresentado no Quadro 4.7, ou seja, a taxa de resposta dos AF's no IFTRAB2004 é de 90,2%.

Com recurso à variável dicotómica ENTRCOMPLET da BD do IFTRAB2004 realizou-se um teste de significância sobre esta taxa de resposta. As categorias desta variável são 1=Sim e 0=Não, pois foi criada para classificar se no agregado obteve-se entrevista completa ou não. Com esta variável as hipóteses estatísticas do teste binomial são as mesmas que foram formuladas anteriormente para a taxa de resposta do IDS2003:

H_0 : A taxa de resposta na população de AF's da Cidade de Maputo é pelo menos 90 por cento ($p \geq 0,90$);

H_1 : A taxa de resposta na população de AF's da Cidade de Maputo é inferior a 90 por cento ($p < 0,90$).

Novamente, para avaliar a significância estatística da taxa de resposta de AF's no IFTRAB2004 na Cidade de Maputo, recorreu-se ao teste binomial implementado no *software* de análise de dados PASW Statistics (v.18; IBM SPSS, Chicago, IL) como descrito em Marôco (2010). Considerou-se o mesmo nível de significância $\alpha=0,05$.

Resultados

Da análise dos dados do IFTRAB2004 feita, a taxa de resposta de AF's é de 90,2%. Portanto, em cerca de 90 por cento dos agregados obteve-se resposta completa. Esta distribuição é apresentada no Gráfico 4.2. O teste estatístico feito, cujos resultados podem ser vistos no Anexo A7, indica que esta taxa de resposta é significativamente igual ou superior a 90 por cento (*valor* $p=0,394 > \alpha=0,05$; $n=1740$).

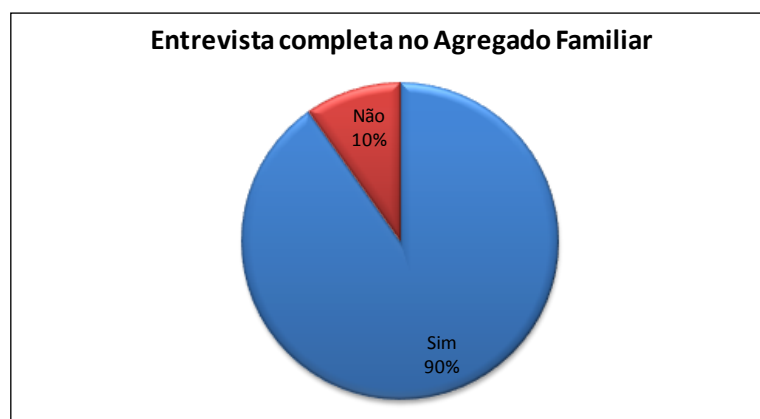


Gráfico 4. 2 – Distribuição percentual da variável ENTRCOMPLET do IFTRAB2004. Fonte: Elaboração própria

Deste modo, os resultados do teste realizado remetem à não rejeição da hipótese nula, pois não há evidências estatisticamente significativas ao nível de 5%. Portanto, a amostra do IFTRAB2004 foi extraída de uma população cuja taxa de resposta é de pelo menos 90 por cento.

Fatores da Não-Resposta no IFTRAB2004

Identicamente ao que foi referido para os não respondentes do IDS2003, a amostra do IFTRAB2004 também pode ser repartida em respondentes ($n_1=1570$) e não respondentes ($n_2=170$). O Quadro 4.8 mostra que dos 170 não respondentes, 61,2% Recusaram-se totalmente em prestar entrevistas e 26,5% não responderam porque todos os membros do AF estiveram Ausentes das suas habitações no período das entrevistas. Ainda, em 8,2% dos 170 agregados houve entrevistas incompletas; em 1,2% dos considerados não respondentes Recusaram-se de prestar toda a entrevista e os restantes 2,9% são agregados que não responderam ao inquérito por Outros motivos.

Fatores de Não-Resposta do AF	Frequência	Porcentagem	Porcentagem acumulada
Recusa total	104	61,2	61,2
Ausência de todo AF	45	26,5	87,6
Entrevista Incompleta	14	8,2	95,9
Recusa durante a entrevista	2	1,2	97,1
Outro	5	2,9	100,0
Total	170	100,0	-

Quadro 4. 8 – Fatores de Não-Resposta no IFTRAB2004

Pode-se associar o anterior Quadro 4.8 de frequências observadas ao outro de frequências esperadas - Quadro 4.9 – para realizar-se um teste de homogeneidade do Qui-quadrado para avaliar a significância das diferenças entre as frequências.

Fatores de Não-Resposta do AF	Frequência Observada	Frequência Esperada	Resíduo
Recusa total	104	34,0	70,0
Ausência de todo AF	45	34,0	11,0
Entrevista Incompleta	14	34,0	-20,0
Recusa durante a entrevista	2	34,0	-32,0
Outro	5	34,0	-29,0
Total	170	-	-

Quadro 4. 9 – Distribuição de frequências esperadas dos não respondentes no IFTRAB2004

As hipóteses do teste são as mesmas formuladas à volta das frequências do Quadro 4.5. Portanto, sob hipótese nula, todos os cinco fatores identificados apresentam igual distribuição de frequências na população de não respondentes do IFTRAB2004.

Resultados do teste

A análise inferencial feita sobre as frequências do Quadro 4.9, indica que as frequências observadas diferem das esperadas ($\chi^2=214,294$; $gl=4$; valor $p=0,000 < \alpha=0,05$, conforme mostra o Anexo A8). Portanto, os fatores da não-resposta no IFTRAB2004 diferem entre si em termos de peso. Assim, com uma confiança de 95% pode-se afirmar que o principal fator é a Recusa total com 61,2% dos casos; em seguida é a Ausência de todos os membros do AF no período das entrevistas (26,5%).

Em terceiro, são as Entrevistas Incompletas (8,2%) seguida de Outros fatores (2,9%). Um fator novo identificado no IFTRAB2004 é a Recusa durante a entrevista com 1,2% dos casos considerados ‘Não-Resposta’ nesta dissertação.

4.1.3. Base de Dados do IOF2009

A amostra do IOF2009 foi desenhada para cobrir um total de 1.200 agregados distribuídos em 100 conglomerados na Cidade de Maputo (conferir no Quadro 4.1). Uma análise da respetiva BD de AF's resulta que foram visitados 1.199 AF's. Estes constam da BD constituída por 240 variáveis. A Figura 4.3 apresenta uma parte dessas variáveis com destaque para RESULTADO, VISITAS e ENTRCOMPLET. Esta última foi criada para classificar os AF em termos de entrevista completa ou não. Portanto, a variável ENTRCOMPLET é dicotómica, seus valores possíveis são 1=Sim ou 0=Não.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Cod_Unico	String	7	0	Codigo do AF	None	None	8	Left	Nominal
2	AE	Numeric	4	0	AE	{1, 0001}...	None	4	Right	Nominal
3	AF	Numeric	3	0	AF	{1, 001}...	None	4	Right	Nominal
4	PR	Numeric	2	0	Provincia	{1, Niassa}...	None	4	Right	Nominal
5	Zone	Numeric	1	0	Regiao	{1, Norte}...	None	5	Right	Nominal
6	DI	Numeric	2	0	DI	{1, 01}...	None	4	Right	Nominal
7	PA	Numeric	1	0	PA	{1, 01}...	None	4	Right	Nominal
8	LO	Numeric	2	0	LO	{1, 01}...	None	4	Right	Nominal
9	AB	Numeric	2	0	AB	{1, 01}...	None	4	Right	Nominal
10	UR	Numeric	1	0	Area de Residencia	{1, Urbana}...	None	5	Right	Nominal
11	dia	Numeric	2	0	dia	{1, 01}...	None	4	Right	Nominal
12	mes	Numeric	2	0	mes	{1, 01}...	None	4	Right	Nominal
13	ano	Numeric	1	0	ano	{1, 2008}...	None	5	Right	Nominal
14	codigo	Numeric	6	0	codigo	None	None	8	Right	Scale
15	RESULTADO	Numeric	6	0	resultado	{1, Compl}...	None	8	Right	Nominal
16	ENTRCOMPLET	Numeric	2	0	Entrevista completa ...	{0, Não}...	None	13	Right	Nominal
17	VISITAS	Numeric	6	0	visitas	None	None	8	Right	Ordinal
18	total	Numeric	6	0	total	None	None	8	Right	Scale
19	religiao	Numeric	11	0	religiao	None	None	8	Right	Nominal
20	controlador	Numeric	6	0	controlador	None	None	8	Right	Scale
21	supervisor	Numeric	6	0	supervisor	None	None	8	Right	Scale
22	revisto_por	Numeric	6	0	revisto_por	None	None	8	Right	Nominal
23	digitado_por	Numeric	6	0	digitado_por	None	None	8	Right	Scale
24	redigitado_por	Numeric	6	0	redigitado_por	None	None	8	Right	Scale

Figura 4. 3 – Uma parte das variáveis da BD de AF's do IOF2009, em SPSS.

Com vista a aumentar a colaboração dos agregados, na recolha de dados do IOF2009 foram efetuadas cinco visitas de insistência principalmente aos agregados ausentes ou de recusa. Estas tentativas resultaram que uns fossem entrevistados na 1ª visita (90,7% dos casos), outros na 2ª (7,2%) e ainda outros na 3ª visita (1,8%). Nas últimas duas visitas foram apurados 0,3% dos agregados planificados no desenho da amostra. O Quadro 4.10 mostra o número de visitas efetuadas e os respetivos agregados abordados.

Número de visitas	Frequência	Percentagem	Percentagem acumulada
1ª visita	1088	90,7	90,7
2ª visita	86	7,2	97,9
3ª visita	22	1,8	99,7
4ª visita	2	0,2	99,9
5ª visita	1	0,1	100,0
Total	1199	100,0	-

Quadro 4. 10 – Frequências de AF's segundo as visitas efetuadas durante o IOF2009. Fonte: Elaboração própria

Contrariamente ao esperado, o número elevado de visitas não foi suficiente para ganhar cooperação de todos os AF. Nisso, foram alcançadas entrevistas completas em 87,7% dos AF e, os restantes, distribuem-se entre a Recusa total (6,2%), a Recusa durante a entrevista (2,2%), a Ausência de todos os membros do agregado durante o período de recolha de dados no conglomerado (2,0%). Os restantes cerca de 2,0% são agregados incapacitados de responder às perguntas do questionário, não responderam à determinadas perguntas ou ainda se encontravam em outras situações que não lhes permitiram participar no IOF2009. A distribuição de frequências é apresentada no Quadro 4.11.

Resultado no Agregado	Frequência	Percentagem	Percentagem acumulada
Completo	1052	87,7	87,7
Recusa total	74	6,2	93,9
Recusa durante a entrevista	26	2,2	96,1
Todo o AF ausente	24	2,0	98,1
AF incapacitado	7	0,6	98,7
Incompleto	5	0,4	99,1
Outro	11	0,9	100,0
Total	1199	100,0	-

Quadro 4. 11 – Resultados da visita ao AF para entrevistas do IOF2009. Fonte: Elaboração própria

Mais uma vez, no cálculo da taxa de resposta são usadas as mesmas notações introduzidas na Subsecção 2.6.1 definidas pela AAPOR (2011). Assim, a taxa de resposta do IOF2009 é obtida com recurso à mesma Equação 2.12.

$$RR_{IOF2009} = \frac{1052}{(1052 + 26) + (24 + 74 + 7 + 5 + 11)} = \frac{1052}{1199} = 0,877$$

Portanto, no IOF2009, a taxa de resposta, sem considerar as entrevistas parciais no numerador, é de 87,7%.

Pode-se testar o quão significativa é esta taxa de resposta sob hipótese nula de que na Cidade de Maputo, a taxa de resposta em inquéritos aos Agregados Familiares é pelo menos 90 por cento. Para isso, foi usado o teste binomial recorrendo-se à variável ENTRCOMPLET e as hipóteses estatísticas por testar são as mesmas formuladas anteriormente tanto para a taxa de resposta do IDS2003 quanto para a do IFTRAB2004:

H_0 : A taxa de resposta na população de AF's da Cidade de Maputo é pelo menos 90 por cento ($p \geq 0,90$);

H_1 : A taxa de resposta na população de AF's da Cidade de Maputo é inferior a 90 por cento ($p < 0,90$).

Portanto, para avaliar a significância estatística da taxa de resposta de AF's no IOF2009 na Cidade de Maputo, recorreu-se ao teste binomial implementado no software PASW Statistics (v.18; IBM SPSS, Chicago, IL) como descrito em Marôco (2010). Considerou-se, novamente, o mesmo nível de significância $\alpha = 0,05$.

Resultados

A amostra dos 1.199 AF pode ser repartida em duas partes: uma dos que consentiram resposta completa (cujas taxas de resposta são 87,7%) e outra dos restantes que, para a presente dissertação são considerados não respondentes. O Gráfico 4.3 mostra esta repartição. Assim, os resultados da análise estatística inferencial, que podem ser conferidos no Anexo A9, indicam que a taxa de resposta é significativamente diferente de 90 por cento, i.é, a taxa de resposta é inferior a 90 por cento ($\text{valor } p = 0,006 < \alpha = 0,05; n = 1199$).



Gráfico 4. 3 – Distribuição percentual da variável ENTRCOMPLET do IOF2009.

Fonte: Elaboração própria

Deste modo, os resultados do teste binomial realizado remetem à rejeição da hipótese nula ao nível de significância de 5%. Portanto, a amostra do IOF2009 foi extraída de uma população cuja taxa de resposta dos AF é inferior a 90 por cento.

Fatores da Não-Resposta no IOF2009

No IOF2009, para os propósitos desta dissertação, a sub amostra dos respondentes é de 1052 AF's. Os restantes 147 são considerados não respondentes. Assim, o Quadro 4.12 mostra a distribuição dos não respondentes segundo os fatores para os quais estes foram assim considerados. Pelo mesmo quadro, pode-se constatar que 50,3% dos não respondentes Recusaram-se totalmente; 17,7% Recusaram-se durante a entrevista e, em 16,3% houve Ausência de todos os membros do AF. Os que não responderam ao inquérito por Outros motivos representam 7,5% dos 147 AF's.

Fatores de Não-Resposta do AF	Frequência	Porcentagem	Porcentagem acumulada
Recusa total	74	50,3	50,3
Recusa durante a entrevista	26	17,7	68,0
Ausência de todo AF	24	16,3	84,4
AF incapacitado	7	4,8	89,1
Entrevista Incompleta	5	3,4	92,5
Outro	11	7,5	100,0
Total	147	100,0	-

Quadro 4. 12 – Fatores de Não-Resposta no IOF2009

Há que comparar o quão significativas são as frequências observadas apresentadas no Quadro 4.12 com as esperadas, sob hipótese de homogeneidade na sua distribuição pelos seis fatores identificados no IOF2009. O Quadro 4.13 mostra a distribuição de frequências esperadas.

Fatores de Não-Resposta do AF	Frequência Observada	Frequência Esperada	Resíduo
Recusa total	74	24,5	49,5
Recusa durante a entrevista	26	24,5	1,5
Ausência de todo AF	24	24,5	-0,5
AF incapacitado	7	24,5	-17,5
Entrevista Incompleta	5	24,5	-19,5
Outro	11	24,5	-13,5
Total	147	-	-

Quadro 4. 13 – Distribuição de frequências esperadas dos não respondentes no IOF2009

Assim, as hipóteses para o teste de Qui-quadrado são:

H_0 : Na população de onde é extraída a amostra as frequências observadas não diferem das esperadas nas respetivas categorias.

H_1 : Na população de onde é extraída a amostra as frequências observadas diferem das esperadas nas respetivas categorias.

Portanto, sob hipótese nula todos os fatores são homogêneos e apresentam igual distribuição de frequências na população de não respondentes do IOF2009.

Resultados do teste

Os resultados da análise inferencial (ver Anexo A10) sobre homogeneidade das frequências do Quadro 4.13, mostram que as frequências observadas diferem das esperadas ($\chi^2=135,571$; $gl=5$; valor $p=0,000 < \alpha=0,05$). Ou seja, os fatores da não-resposta no IOF2009 são seis e diferentes entre si. O principal fator é a Recusa total (50,3%), em segundo vêm a Recusa durante a entrevista (17,7%) e a Ausência dos AF's (16,3%)²⁶. Em terceiro seguem-se Outros fatores (7,5%) e, depois os restantes fatores: Incapacidade dos membros do AF (4,8%) e Entrevistas incompletas (3,4%).

²⁶ Para estes dois fatores as diferenças não são estatisticamente significativas ($\chi^2=0,080$; $gl=1$; valor $p=0,777 > \alpha=0,05$), conforme se pode conferir no Anexo A10.1.

4.2. ANÁLISE COMPARATIVA DOS DADOS

As BD's dos três inquéritos mencionados foram unidas numa única com vista a permitir análise comparativa de variáveis. A Figura 4.4 mostra uma parte das variáveis da BD conjunta. Por exemplo, a variável que aparece selecionada 'INQUERITO' foi criada como forma de identificar casos que fazem parte de um ou outro inquérito, ou seja, as categorias desta variável são 1=IDS2003, 2=IFTRAB2004 e 3=IOF2009.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	PROVIN	Numeric	2	0	Provincia	{1, Niassa}...	None	7	Right	Nominal
2	ID	Numeric	3	0	Numero do con...	None	None	7	Right	Scale
3	AF	Numeric	3	0	Numero do agr...	None	None	6	Right	Scale
4	Agregado	Numeric	3	0	Filtrar Agregad...	None	None	8	Right	Scale
5	INQUIRIDOR	Numeric	3	0	Numero Inquirid...	None	None	7	Right	Scale
6	RESULTADO	Numeric	2	0	Resultado	{1, Complet...	None	7	Right	Nominal
7	ENTRCOMPLET	Numeric	2	0	Entrevista com...	{1, Sim}...	None	5	Right	Nominal
8	VISITAS	Numeric	1	0	Número total d...	{1, 1ª Visita}...	None	7	Right	Ordinal
9	SUPERVISOR	Numeric	3	0	Supervisor	None	None	7	Right	Scale
10	CONTROLADOR	Numeric	3	0	Controlador	None	None	6	Right	Scale
11	INQUERITO	Numeric	2	0	Tipo de inquérito	{1, IDS2003...	None	6	Right	Nominal

Figura 4. 4 – Uma parte das variáveis da BD conjunta dos três inquéritos, em SPSS

Assim, em termos comparativos o Quadro 4.14 mostra que no IOF2009 foram efetuadas mais visitas aos AF's do que no IDS2003. Mais de três quartos dos agregados foram entrevistados na 1ª visita nos três inquéritos.

Número de visitas	IDS2003		IFTRAB2004		IOF2009	
	Frequência	Percentagem	Frequência	Percentagem	Frequência	Percentagem
1ª Visita	707	75,1	1587	91,2	1088	90,7
2ª Visita	96	10,2	83	4,8	86	7,2
3ª Visita	138	14,7	68	3,9	22	1,8
4ª Visita	0	0,0	2	0,1	2	0,2
5ª Visita	0	0,0	0	0,0	1	0,1
Total	941	100,0	1740	100,0	1199	100,0

Quadro 4. 14 – Agregados entrevistados segundo o número de visitas

O Quadro 4.14 mostra que o número de entrevistas esperadas pode ser efetivamente alcançado com o aumento do número de visitas de insistência. As visitas de insistências visam aumentar a colaboração dos AF's selecionados. Conforme aconselharam os autores Biemer e Lyberg em 2003, para reduzir a contribuição dos casos de não-contacto na taxa de resposta e o consequente enviesamento da não-

resposta, a principal estratégia pode ser a persistência, continuar a tentar o contacto com a unidade amostral até que finalmente se consiga a entrevista ou tenha que se desistir.

Importa avaliar a relação de associação entre o tipo de inquérito e o resultado obtido na entrevista. O Quadro 4.15 mostra a distribuição de frequências do resultado da entrevista em termos de fatores da não-resposta – variável RESULTADO – e o tipo de inquérito. De forma simples, para os objetivos desta dissertação, interessa avaliar se os fatores de não-resposta dependem ou não do tipo de inquérito.

Fator da Não-Resposta	IDS2003		IFTRAB2004		IOF2009	
	Frequência	Percentagem	Frequência	Percentagem	Frequência	Percentagem
Todo AF Ausente	171	81,0	45	26,5	24	16,3
Recusa Total	18	8,5	104	61,2	74	50,3
Recusa Durante a entrevista	0	0,0	2	1,2	26	17,7
Entrevista Incompleta	3	1,4	14	8,2	5	3,4
AF Incapacitado	14	6,6	0	0,0	7	4,8
Outro	5	2,4	5	2,9	11	7,5
Total	211	100	170	100	147	100

Quadro 4. 15 – Casos de Não-Resposta por inquérito segundo seus fatores

Pelo Quadro 4.15, dos casos considerados não respondentes, nenhum deles foi assim classificado por ter-se Recusado Durante a Entrevista do IDS2003; de igual modo, nenhum caso foi registado de membros do AF na condição de Incapacitados em responder aos questionários durante a recolha de dados do IFTRAB2004.

Todavia, pelo facto de se ter amostras aleatórias, nos seis fatores identificados sob hipótese de independência entre os fatores da não-resposta e o tipo de inquérito conduzido, as frequências observadas não devem ser estatisticamente diferentes das esperadas. O Quadro 4.16 mostra a distribuição de frequências que resultam do cruzamento das variáveis ‘Tipo de Inquérito’ e ‘Fatores da Não-Resposta’.

		Frequências	Tipo de inquérito		
			IDS2003	IFTRAB2004	IOF2009
Fatores da Não-Resposta	Todo AF Ausente	Observadas	171	45	24
		Esperadas	95,9	77,3	66,8
	Recusa Total	Observadas	18	104	74
		Esperadas	78,3	63,1	54,6
	Recusa Durante a entrevista	Observadas	0	2	26
		Esperadas	11,2	9,0	7,8
	Entrevista Incompleta	Observadas	3	14	5
		Esperadas	8,8	7,1	6,1
	AF Incapacitado	Observadas	14	0	7
		Esperadas	8,4	6,8	5,8
	Outro	Observadas	5	5	11
		Esperadas	8,4	6,8	5,8
	Total	Observadas	211	170	147
		Esperadas	211,0	170,0	147,0

Quadro 4. 16 – Distribuição de frequências para teste de independência do Qui-quadrado

Conforme referenciado na Subsecção 3.5.2, segundo Marôco (2010), o teste Qui-quadrado permite testar se a distribuição dos elementos da amostra global é aleatória ou não. Se a distribuição dos elementos observados for aleatória, isso é equivalente a afirmar que as distribuições das categorias das variáveis consideradas são independentes das classes de uma segunda variável. No caso desta dissertação, o teste foi realizado sob as seguintes hipóteses estatísticas:

H_0 : *Os fatores da não-resposta são independentes do tipo de inquérito realizado na Cidade de Maputo.*

H_1 : *Os fatores da não-resposta não são independentes do tipo de inquérito realizado na Cidade de Maputo.*

Resultados

Em alusão às frequências observadas e esperadas apresentadas no Quadro 4.16 e da análise estatística inferencial cujos resultados constam do Anexo A11 – segundo Marôco (2010) – segue que as diferenças entre estas frequências são estatisticamente significativas ($\chi^2=266,636$; $gl=10$; $valor\ p=0,000 < \alpha=0,05$). Ou de forma equivalente, com uma confiança de 95% pode-se afirmar que os fatores de não-resposta não são independentes do tipo de inquérito realizado na Cidade de Maputo, ou seja, inquéritos com objetivos e conteúdos diferentes podem produzir fatores de não-resposta também diferentes.

Adicionalmente, o grau de associação entre as variáveis ‘Fatores da Não-Resposta’ e ‘Tipo de Inquérito’ é significativamente elevado, conforme mostram as

respetivas medidas de associação cujos detalhes constam do Anexo A11 ($\Phi=0,711$; V de Cramer= $0,502$; Coeficiente de Contingência= $0,579$; valor $p=0,000 < \alpha=0,05$).

4.3. DISCUSSÃO DOS RESULTADOS

Esta secção apresenta a discussão dos principais resultados obtidos da análise de dados. Um breve enquadramento a partir da revisão bibliográfica é apresentado ao longo da discussão.

4.3.1. As Taxas de Resposta por inquérito

Os resultados obtidos da análise precedente, atendendo ao facto de os desenhos amostrais dos três inquéritos serem semelhantes, podem ser comparáveis, particularmente no que se refere às taxas de resposta, conforme refere AAPOR (2011). Assim, as análises estatísticas permitem afirmar que a taxa de resposta de 77,6% no IDS2003 está significativamente abaixo dos 90%. Este resultado remete-nos a afirmar que na Cidade de Maputo onde foi extraída a amostra do IDS2003 a taxa de resposta dos AF's está abaixo da esperada.

Esta taxa é influenciada pelos erros não amostrais de não-resposta. A análise revelou que os fatores que influenciaram esta taxa de resposta são, principalmente, a Ausência de todos os membros do agregado no período das entrevistas, a recusa total e a existência de AF's selecionados cujos membros foram incapazes de responder ao questionário. Alguns destes fatores de não-resposta têm sido levantados por pesquisadores. Com efeito, Biemer e Lyberg (2003) apontam uma série de fatores e efeitos da não-resposta, nomeadamente, a fraca motivação dos respondentes, a falta de tempo por parte dos respondentes, a indisponibilidade devido a doença, a ocupação laboral a que são sujeitos os respondentes, os problemas ligados à língua usada e ao período concebido para a recolha de dados. E, estes autores salientam que tais fatores concorrem para as recusas ou ausência dos potenciais respondentes principalmente em inquéritos domiciliários

No que se refere ao IFTRAB2004, os resultados mostram que neste inquérito, que se realizou um ano depois do IDS2003, a taxa de resposta de 90,2% não é

estatisticamente diferente da taxa declarada no desenho da amostra. Portanto, a amostra do IFTRAB2004 foi extraída de uma população cuja taxa de resposta é de pelo menos 90%. As mensagens de sensibilização veiculadas durante o decurso do trabalho de campo, o longo período de recolha de dados de 12 meses, as visitas de insistências aos AF's, estes e outros fatores provavelmente tenham contribuído para a subida desta taxa comparativamente à do IDS2003.

Há que recordar que no IDS2003 a recolha de dados foi de 4 meses. Conforme referem Biemer e Lyberg (2003), algumas ações podem ser tomadas com vista a aumentar a colaboração dos respondentes. Por exemplo, a classificação das não-respostas a partir dos seus fatores pode ajudar o processo de recolha de dados especialmente aos respondentes que se ausentam das suas habitações no período de entrevistas. Para esses casos, segundo Biemer e Lyberg, os respondentes de difícil contacto podem ser abordados usando-se procedimentos eficazes de rastreamento e ainda usando padrões eficientes para tentativas de contacto repetidas.

Assim, os resultados apresentados nas Secções 4.1 e 4.2 mostram que os fatores que estão associados a não-resposta no IFTRAB2004 são as recusas dos AF's, as ausência dos AF's no período das entrevistas, as entrevistas incompletas. Alguns entrevistados que se dispuseram em responder, no decurso da entrevista recusaram-se em responder a totalidade da entrevista. Este último fator pode estar associado com o que salientam Groves *et al.* (1992) citados por Biemer e Lyberg (2003): o volume do trabalho dos inquiridores, a disposição e as características dos inquiridores, a extensão do questionário, a duração das entrevistas, o conteúdo das perguntas e o desenho do próprio questionário exercem influência na cooperação dos respondentes.

O IOF2009 cujo trabalho de recolha de dados durou 12 meses teve uma taxa de resposta de 87,7%. Esta taxa é significativamente baixa atendendo que na população de onde se extraía a amostra a taxa de resposta mínima é 90%. Os fatores que influenciaram esta taxa são praticamente os mesmos reportados para o IDS2003 e IFTRAB2004, nomeadamente, a recusa total, a recusa durante a entrevista, a ausência de todos os membros do AF no período de entrevistas e outros fatores.

Os resultados aqui apresentados mostram que as taxas de resposta dos inquiridos aos AF's na Cidade de Maputo são estatisticamente diferentes e cada vez

menores do que as esperadas. Foi referenciado por Biemer e Lyberg (2003) que para avaliação da qualidade dos resultados de um inquérito por amostragem é importante ter em conta dois determinantes: a taxa de resposta e o enviesamento resultante da não-resposta. Ainda, conforme referem estes autores, muito mais esforço deve ser desenvolvido pelos pesquisadores com vista a aumentar o mais possível a taxa de resposta e, conseqüentemente, minimizar os riscos do enviesamento da não-resposta.

Quanto aos fatores da não-resposta que influenciam as taxas de resposta, da análise feita às três BD's foram identificados um total de seis fatores aos inquéritos na Cidade de Maputo: Ausência de todos os membros do AF, Recusa total em participar no inquérito, a Recusa durante a entrevista, as Entrevistas incompletas, a Incapacidade dos membros do AF em prestar informação e Outros fatores (falecimentos no período de entrevista, língua usada).

Os principais fatores da não-resposta são fundamentalmente as ausências dos agregados familiares e as recusas em prestar entrevistas. Na sua obra *Survey Sampling*, Kish (1965) salienta a necessidade de os pesquisadores identificarem os fatores da não-resposta e de outros erros não amostrais para o seu controlo e redução, prevê-los em futuras pesquisas e estimar os seus possíveis efeitos nos resultados dos inquéritos. Alguns dos fatores aqui identificados também foram referenciados por Kish (1965) aos que chamou de: (1) *Não em casa*, (2) *Recusas* e (3) *Incapacidade* ou *Inaptidão* do respondente por doença, problemas psíquicos ou mentais. As barreiras criadas pela língua usada ou pelo analfabetismo são, segundo Kish, inclusas nesta categoria.

4.3.2. Os Fatores de Não-Resposta por inquérito

Na análise dos resultados desta dissertação, foi encontrada alguma relação estatisticamente significativa entre os fatores da não-resposta e o tipo de inquérito. O IDS2003 teve seus conteúdos principais ligados à saúde dos membros dos agregados familiares e sua taxa de resposta é a mais baixa comparativamente à do IFTRA2004 e do IOF2009. As mensagens sobre os aspetos ligados à saúde das pessoas provavelmente não tenham sido veiculadas e recebidas de forma adequada no seio da população da Cidade de Maputo. Um dos temas abordados num dos questionários do IDS2003 foi a problemática do VIH/SIDA e, a cooperação pode ter baixado porque as pessoas se sentem inseguras em falar de assuntos relacionados com esta pandemia.

Outra causa da baixa taxa de resposta no IDS2003 provavelmente tenha a ver com os instrumentos antropométricos que as brigadas de recolha de dados levavam consigo, pois as pessoas não sabiam do que se tratava e, por escassez de informação, a colaboração não foi a desejável.

Um ano depois do IDS2003 realizou-se o IFTRAB2004 que fundamentalmente se concentrou em recolher informação sobre a força de trabalho, o emprego, o desemprego e as características da mão-de-obra moçambicana. Neste inquérito a taxa de resposta excedeu significativamente os 90%. Esta subida de taxa de resposta pode estar relacionada com o elevado desemprego²⁷ na Cidade de Maputo e provavelmente a cooperação tida dos respondentes tenha sido espécie de *“Eu colaboro respondendo e depois procuro oportunidades de emprego junto das brigadas de recolha de dados, pois trata-se de inquérito sobre trabalho”*.

Enquanto que no IOF2009 cujo objetivo principal foi recolher dados sobre receitas e despesas das pessoas e dos agregados familiares, a taxa de resposta voltou ao decréscimo significativo. Essa descida de taxa de resposta provavelmente esteja ligada ao facto de as pessoas se mostrarem receosas em declarar seus rendimentos e os dos seus AF's. Uma forma encontrada pelos potenciais respondentes foi de não participar no IOF2009 recusando totalmente ou pautando por se ausentar das suas habitações no período de entrevistas. Ainda conforme refere Kish (1965), a natureza e conteúdo das perguntas podem afetar a participação dos respondentes nos inquéritos, pois em diversas culturas, assuntos como receitas, despesas, religião, sexo, política podem ser de algum modo fáceis ou difíceis discutir. E, Kish acrescenta que os efeitos desses vários fatores nas recusas não aparecem de forma isolada pois interagem entre si. Os níveis de significância dos fatores identificados nesta dissertação por inquérito são apresentados no Anexo A12.

²⁷ O Anexo A4 apresenta as taxas de desemprego derivadas do IAF2002/03 e do IFTRAB2004.

5. CONCLUSÕES E RECOMENDAÇÕES

Neste capítulo são apresentadas algumas conclusões consideradas principais, recomendações e proposta de estudos futuros.

5.1. PRINCIPAIS CONCLUSÕES

O objetivo desta dissertação foi de descrever os erros não amostrais e identificar os principais fatores dos erros de não-resposta em inquéritos por amostragem realizados na Cidade de Maputo. Especificamente, foi interesse desta dissertação descrever os tipos de erros não amostrais, calcular as taxas de resposta no IDS2003, IFTRAB2004, IOF2009, três inquéritos nacionais realizados pelo INE-M para a Cidade de Maputo e, destes inquéritos, identificar os principais fatores que influenciam a não-resposta.

5.1.1. Fontes de Erros Não Amostrais

Da literatura sobre o tema em estudo há que tecer algumas conclusões principais. Os inquéritos por amostragem podem ser conduzidos sob um plano de sondagem probabilístico ou não probabilístico. Independentemente de se seguir um plano ou outro há sempre erros associados aos resultados cuja minimização da sua ocorrência é crucial. E, os erros associados aos resultados do inquérito podem ser amostrais (ligados a amostragem em si) e não amostrais (alheios ao processo de amostragem).

Esta dissertação concentrou sua atenção aos erros não amostrais. Os erros não amostrais devem-se às diferentes etapas em que percorre o inquérito e podem ser sistematizados, segundo Cochran (1977), Kish (1965), Biemer e Lyberg (2003) em quatro categorias (erros de cobertura, de medição ou especificação, de processamento e de não-resposta). Porém, Biemer e Lyberg (2003), na sua obra *Introduction to Survey*

Quality, são mais específicos e distinguem os erros não amostrais a partir das suas cinco fontes e respetivos subtipos:

- (1) Erros de especificação (dos conceitos, dos objetivos, dados a recolher);
- (2) Erros da Base de Sondagem (de omissão, inclusão errada, duplicação);
- (3) Erros de Não-resposta (toda unidade, parte da unidade, do item específico, resposta inadequada ou incompleta);
- (4) Erros de Medida (dos instrumentos de recolha de dados, do entrevistador, do respondente, da forma de recolha de dados, do sistema de informação, do ambiente de recolha de dados);
- (5) Erros de Processamento (de edição, de codificação, de digitação, de ponderação, de tabulação).

5.1.2. Taxa de Resposta e Principais Fatores da não-resposta

Deste estudo, pode-se concluir que a taxa de resposta aos inquéritos aos agregados familiares variam de acordo com o tipo de inquérito e, os seus valores em geral tendem a ser abaixo do esperado. As amostras são desenhadas assumindo-se que na população de agregados familiares a taxa de resposta é de pelo menos 90%. Mas os resultados sugerem que em Inquéritos Demográficos e de Saúde a taxa de resposta é abaixo dos 90%. Em Inquéritos à Força do Trabalho a taxa tende a atingir significativamente os 90% mas não muito superior a este valor (Gráfico 5.1). Estudos pormenorizados podem ser realizados com vista a apurar as motivações de significativa participação dos agregados neste tipo de inquérito. Provavelmente, a cooperação dos agregados seja induzida pela natureza deste tipo de inquérito que está ligada ao trabalho atendendo à elevada taxa de desemprego na Cidade de Maputo.

Os resultados que sustentam as conclusões apresentadas nesta dissertação suscitam alguma preocupação quanto à tendência de decaimento da taxa de resposta. Aliás a este respeito, Steeh em 1981 mostrou sua preocupação e demonstrou, num estudo, que havia um crescimento substancial das não-respostas devido,

principalmente, à recusa dos entrevistados. Os resultados do estudo de Steeh demonstraram ainda que estas tendências estão relacionadas com o nível de urbanização, pois quanto mais alto for o nível de urbanização, as taxas de recusa tendem a ser mais altas. Portanto, atendendo ao crescente grau de urbanização que caracteriza a Cidade de Maputo nos últimos anos, há toda a preocupação de encontrar formas de obter maior colaboração dos potenciais respondentes aos inquéritos.

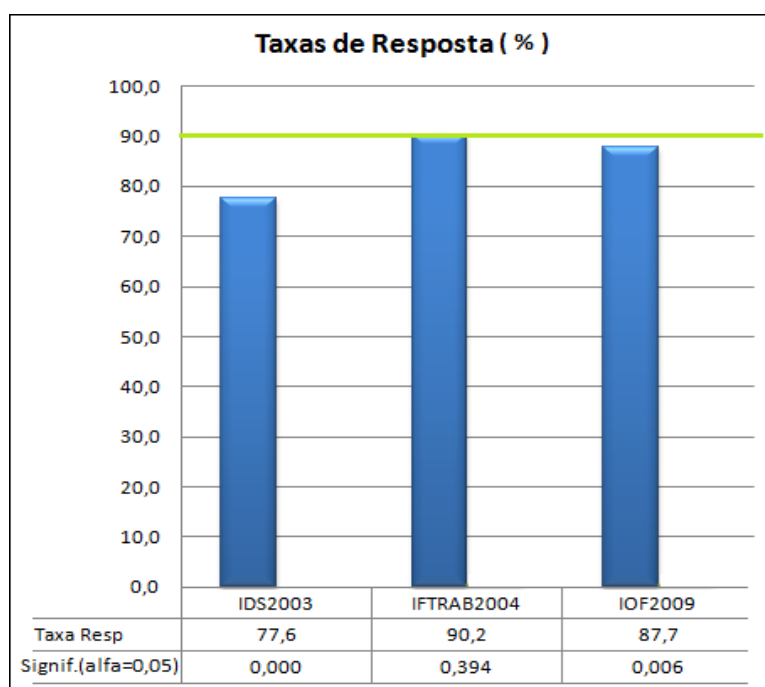


Gráfico 5. 1 – Taxas de resposta por inquérito e sua significância

Ainda desta dissertação, com um grau de confiança de 95%, pode-se concluir que no inquérito ao orçamento familiar a taxa é significativamente baixa do que a esperada. De igual modo, com maior profundidade poder-se-á apurar, em estudos futuros e essencialmente com esse objetivo, as reais causas da fraca colaboração dos agregados familiares em inquéritos desta natureza. A literatura defende que em diversas culturas, assuntos como receitas, despesas, religião, sexo, política podem ser de algum modo fáceis ou difíceis discutir no seio dos respondentes (Kish, 1965). E, Kish acrescenta que os efeitos desses vários fatores nas recusas não aparecem de forma isolada pois interagem entre si.

A taxa de resposta é influenciada por diversos fatores ligados a não-resposta (Biemer e Lyberg, 2003). Os fatores identificados nesta dissertação que concorrem para redução da taxa de resposta aos inquéritos na Cidade de Maputo são as recusas totais, as ausências dos respondentes, a incapacidade ou inaptidão dos respondentes, as recusas durante a entrevista, as entrevistas incompletas e outros fatores ligados à infelicidade, barreiras linguísticas (Gráfico 5.2). Estes fatores são concordantes com os previstos na literatura.

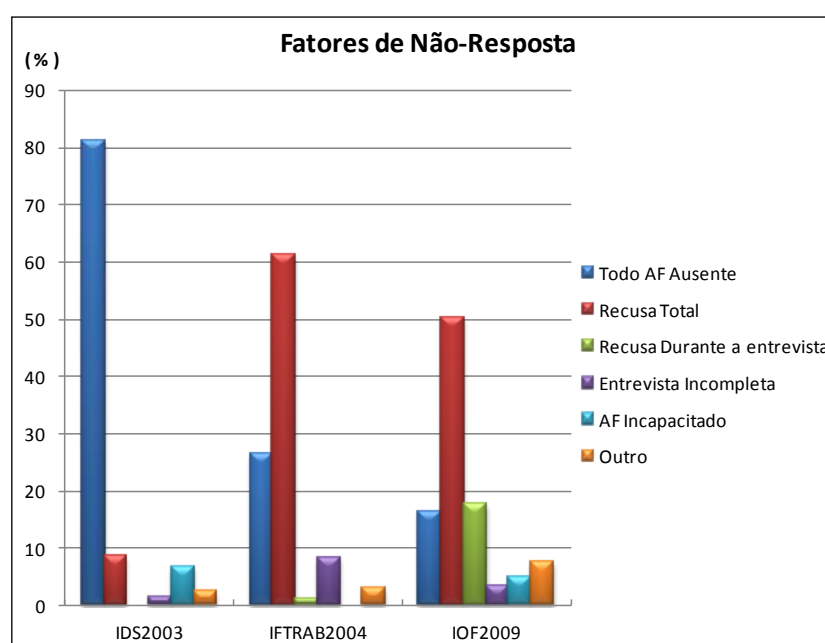


Gráfico 5. 2 – Fatores de Não-Resposta por inquérito

Destes fatores, são considerados os principais as recusas e as ausências dos agregados das suas habitações no período de entrevistas.

5.2. RECOMENDAÇÕES

Como resultado da presente dissertação, para garantir melhor precisão das estimativas derivadas de um inquérito na Cidade de Maputo, várias ações podem ser tomadas conducentes ao controlo e prevenção de erros de não-resposta. Dentre elas, são apresentadas as seguintes.

5.2.1. A nível institucional

O INE-M e outras instituições de pesquisa podem melhorar os níveis de cooperação dos respondentes da Cidade de Maputo adotando várias medidas das quais se salientam: o trabalho do inquérito, os seus objetivos, a duração do processo de recolha de dados e o período das entrevistas devem ser divulgados continuamente; devem ser difundidas a legitimidade do inquérito e das instituições envolvidas; devem ser garantidas várias tentativas de contacto com os respondentes; deve-se garantir o anonimato e a confidencialidade das respostas; os inquiridores devem ser identificados tanto pela documentação quanto à indumentária; fornecer/divulgar-se contactos telefónicos para esclarecimentos; despertar e motivar o interesse dos respondentes com mensagens convidativas à cooperação salientando a importância do inquérito; a capacitação de inquiridores deve ser uniforme, simultânea e num mesmo local; os inquiridores devem ser capacitados para conduzir entrevistas em línguas mais faladas na Cidade de Maputo inclusive a inglesa.

5.2.2. Modo e condições definidas do inquérito

A natureza do conteúdo pesquisado deve ser do conhecimento de todo pessoal envolvido no inquérito e dos respondentes; os métodos de recrutamento e seleção de inquiridores devem ser objetivos e claros; os respondentes à entrevista no domicílio devem ser definidos claramente; os critérios de substituição²⁸ dos não respondentes devem ser definidos e claros; o tipo de entrevista (domiciliária) e face-a-face devem ser do conhecimento dos respondentes; as condições e o ambiente da entrevista devem ser tais que os respondentes sintam alguma privacidade.

²⁸ As substituições de unidades amostrais são sugeridas como remédio para os não respondentes. Entretanto, Kish (1965) considera que esta prática é errada porque os substitutos assemelham-se aos respondentes ao invés dos não respondentes.

5.2.3. A nível do questionário usado

No que se refere ao desenho do questionário, como forma de elevar a taxa de resposta, dentre várias medidas sugerem-se as seguintes: reduzir a extensão do questionário com a inclusão de perguntas essenciais; o questionário deve ser o mais curto e simples possível em função do pesquisado; as perguntas do questionário devem ser em linguagem mais simples e acessível; evitar-se o uso de palavras difíceis na formulação de perguntas do questionário; as perguntas devem ser formuladas sem ambiguidades; a ordem em que são arrumadas as perguntas é crucial para evitar-se a não-resposta; durante a entrevista, advertir aos respondentes sobre questões sensíveis do questionário, mas apelando da necessidade de resposta; evitar a inserção/supressão de perguntas ao questionário durante a recolha de dados; as perguntas devem ser tais que meçam exatamente o pesquisado independentemente da língua usada na entrevista; os inquiridores devem ser suficientemente treinados de modo que, durante as entrevistas interpessoais, compreendam a intenção do pesquisador em cada pergunta do questionário; reduzir a carga sobre os respondentes por fadiga e desatenção.

5.2.4. A nível dos inquiridores vs respondentes

No recrutamento de inquiridores algumas características, tais como idade, género, raça, classe social, nível de educação do inquiridores devem ser tidas em conta, pois estas influenciam a cooperação dos respondentes. Devem ser recrutados inquiridores que mostrem elevadas habilidades no contacto com os respondentes e com maior poder de persuasão; os inquiridores devem mostrar um alto espírito organizacional; recrutar inquiridores cujas características e aparências se assemelhem às da população pesquisada em conformidade aos objetivos do inquérito; a capacidade e a experiência dos inquiridores deve ser tida em conta, pois isto lhes torna hábeis em obter maior cooperação dos respondentes; os inquiridores mais confiantes e habilidosos tendem a obter altas taxas de cooperação dos respondentes; os inquiridores devem transmitir confiança e domínio dos conteúdos do inquérito; a

atitude e o comportamento do inquiridor devem ser os mais recomendados possíveis no contacto com os respondentes.

5.3. LIMITAÇÕES E PROPOSTA DE ESTUDOS FUTUROS

A elaboração desta dissertação enfrentou dificuldades quanto ao acesso à informação pertinente, principalmente às bases de dados do INE-M, aos documentos metodológicos que orientaram cada um dos três inquéritos aqui focalizados. Outra dificuldade prendeu-se com o acesso via Web do INE-M da informação auxiliar resultante dos censos populacionais de 1997 e 2007 e de relatórios disseminados. Uma parte de informação requerida não está disponível no *site* do INE-M e, quando é disponível o seu acesso é difícil provavelmente pelo tamanho elevado dos ficheiros. Estas dificuldades, de algum modo, concorreram para que os dados analisados nesta dissertação e o cruzamento de informação fossem limitados.

Outra limitação prende-se com o facto de ter-se analisado informação previamente recolhida pelo INE-M. Os questionários utilizados não incluíram variáveis suficientes para apurar todos os casos e fatores de não-resposta. O autor desta dissertação limitou-se em analisar as variáveis ora determinadas pelo INE-M. A recolha de dados primários pelo autor desta dissertação daria mais subsídios aos resultados aqui apresentados. Outra limitação, ainda, está ligada ao cálculo das taxas de resposta. Estas foram calculadas usando-se dados não ponderados, pois as BD's analisadas não incluem fatores de extrapolação. Quando as probabilidades de seleção variam consideravelmente de uma unidade para outra, as taxas de resposta ponderadas e não-ponderadas podem ser muito diferentes (Biemer e Lyberg, 2003).

Apesar das limitações referidas anteriormente, surgiram questões que não foram respondidas no âmbito desta dissertação mas cuja pertinência se julga interessante para os produtores de informação estatística em estudos futuros:

- 1- Quais são as reais causas de ausências ou de recusas dos potenciais respondentes aos estudos na Cidade de Maputo. Um estudo desta natureza abrangendo a população de não respondentes poderia apoiar a traçar medidas com vista a incrementar a cooperação e elevar a taxa de resposta. Ainda, poder-se-ia segmentar a população de não respondentes por

caraterísticas socio-demográficas e ou estratos geográficos ou por outras variáveis de interesse.

- 2- Outro estudo considerado interessante seria sobre o impacto da não-resposta no enviesamento dos resultados seguindo a abordagem sugerida, principalmente, por Biemer e Lyberg (2003:68-80, 91-115) em variáveis selecionadas nos três inquéritos para a Cidade de Maputo. Esse estudo podia trazer medidas quantitativas da qualidade dos resultados face às diferentes taxas de resposta em função da magnitude do enviesamento.

6. BIBLIOGRAFIA E REFERÊNCIAS

- Aleixo, S. et al (2011). *Erros Não Amostrais – Uma Floresta de Enganos. Boletim da Sociedade Portuguesa de Estatística*. Publicação semestral. Primavera de 2011. Versão electrónica disponível no sítio <http://www.spestatistica.pt>, acedido em 22 de Março de 2012
- Assael, H. e Keon, J. (1982). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing* Vol. 46, No. 2 (Spring, 1982), pp. 114-123 Artigo disponível no URL: <http://www.jstor.org/stable/3203346>, acedido em 13 de Maio de 2012
- Aubyn, A. e Venes, N. (2011). *Análise de dados. Aplicação às ciências económicas e empresariais*. Verlag Dashöfer. Lisboa
- Bassi, F. e Fabbris, L. (1997). Estimators of Nonsampling Errors in Interview – Reinterview Supervised Surveys with Interpenetrated Assignments. In *Survey Measurement and Process Quality*, eds. Lars Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. New York: Wiley, 733–751.
- Barriball, K. L., & While, A. E. (1999). Non-response in survey research: a methodological discussion and development of an explanatory model. *Journal of Advanced Nursing*, 677-686. DOI:10.1046/j.1365-2648.1999.01117.x
- Biemer, P. P e Lyberg, L. E. (2003). *Introduction to Survey Quality*. John Wiley & Sons, New York
- Cochran, W. G. (1977). *Sampling Techniques*. 3rd Edition. John Wiley & Sons, New York
- Deutskens, E., Ruyter, K. d., Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters* 15:1, 21–36. DOI:10.1023/B:MARK.0000021968.86465.00

- Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research* 38, 1–18. DOI:10.1016/j.ssresearch.2008.03.007
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior* 26, 132–139. DOI:10.1016/j.chb.2009.10.015
- Freitas, A. (2002). Caderno de leitura e apoio ao curso de Amostragem para Escola de Verão de Maputo (Setembro 2002). Lisboa
- Fricker, S.; Tourangeau, R. (2010). Examining the Relationship between Nonresponse propensity and Data Quality in two National Household Surveys. *Public Opinion Quarterly*, Vol. 74, Nº 5, 2010, pp. 934-955. Artigo disponível em: <http://poq.oxfordjournals.org/content/70/5/646.short> acedido em 15 de Maio de 2012
- Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, Vol. 70, Nº 5, Special Issue 2006, 646–675. DOI:10.1093 / poq / nfl033
- Groves, R. M. e Lyberg L. (2010). Total Survey errors. Past, Present, and Future. *Public Opinion Quarterly*, Vol. 74, Nº 5, 849-879. Artigo disponível em: <http://poq.oxfordjournals.org/content/70/5/646.short> acedido em 14 de Maio de 2012
- Groves, R. M. e Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. A Meta-Analysis. *Public Opinion Quarterly*, Vol. 72, Nº 2, 167–189. DOI:10.1093/poq/nfn011
- Guimarães, R. C. e Cabral, J. S. (2011). Estatística. 2ª Edição. Verlag Dashöfer. Lisboa
- Instituto Nacional de Estatística de Moçambique (1997). Resultados definitivos do Recenseamento Geral da População e Habitação de 1997 – IIRGPH 1997. Instituto Nacional de Estatística, Maputo, Moçambique

- Instituto Nacional de Estatística de Moçambique (2011). Indicadores Básicos de Moçambique. Instituto Nacional de Estatística, Maputo, Moçambique. Versão eletrónica disponível no sítio: <http://www.ine.gov.mz/Links.aspx>
- Instituto Nacional de Estatística de Moçambique (2007). Resultados definitivos do Recenseamento Geral da População e Habitação de 2007 – IIIRGPH 2007. Instituto Nacional de Estatística, Maputo, Moçambique
- Instituto Nacional de Estatística de Moçambique (2003). Recomendações metodológicas do desenho do Inquérito Demográfico e de Saúde de 2003 – IDS 2003. Instituto Nacional de Estatística, Maputo, Moçambique
- Instituto Nacional de Estatística de Moçambique (2003). Manual de Inquiridor do Inquérito Demográfico e de Saúde de 2003 – IDS 2003. Instituto Nacional de Estatística, Maputo, Moçambique
- Instituto Nacional de Estatística de Moçambique (2004). Manual do Inquiridor do Inquérito Integrado à Força do Trabalho de 2004/2005 – IFTRAB 2004. Instituto Nacional de Estatística, Maputo, Moçambique
- Instituto Nacional de Estatística de Moçambique (2009). Manual de Inquiridor do Inquérito ao Orçamento Familiar de 2009 – IOF 2009. Instituto Nacional de Estatística, Maputo, Moçambique
- James, K. M.; Ziegenfuss, J. Y.; Tilburt, J. C.; Harris, A. M.; e Beebe, T. J. (2011). Getting Physicians to Respond: The Impact of Incentive Type and Timing on Physician Survey Response Rates. Health Research and Educational Trust. Vol. 46, Nº 1, 232-242. DOI: 10.1111/j.1475-6773.2010.01181.x
- Kish, L. (1965). Survey Sampling. Wiley Classics Library. John Wiley & Sons, New York
- Kreuter, F., Muller, G. e Trappmann, M. (2010). Nonresponse and measurement error in employment research. Making use of administrative data. *Public Opinion Quarterly*, Vol. 74, No. 5, 2010, pp. 880-906. Artigo disponível em: <http://pog.oxfordjournals.org/content/70/5/646.short> acessado em 13 de Maio de 2012

- Langer, G. (2003). About Response Rates. Some Unresolved Questions. Public Perspective, May/June 2003, 16 - 18. Disponível no sítio: <http://www.aapor.org> acedido em 24 de Julho de 2012
- Lyberg, L. et al (1997). Survey Measurement and Process Quality. Wiley. New York,
- Marôco, J. (2010). Análise Estatística com o PASW Statistics (ex-SPSS). Report Number. Pêro Pinheiro
- Megill, D. (2002). Recomendação para o desenho da amostra do inquérito de agregados familiares 2002 de Moçambique – IAF 2002/3. Instituto Nacional de Estatística, Maputo, Moçambique
- Megill, D. (2008). Sampling Recommendations for the Mozambique Inquérito ao Orçamento Familiar - IOF2009. Instituto Nacional de Estatística, Maputo, Moçambique
- Porter, S. R., e Whitcomb, M. E. (2003). The Impact of Lottery Incentives on Student Survey Response Rates. Research In Higher Education, Vol. 44, Nº 4, 389-404
- Reis, F. L. (2010). Como elaborar uma dissertação de Mestrado. Segundo Bolonha. Guia prático, 2ª edição, PACTOR, Lisboa
- Sakshaug, J. W., Yan, T., Tourangeau, R. (2010). Nonresponse error, Measurement error, and mode of data collection. Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*. Vol. 74, No. 5, 2010, 907-933. Artigo disponível em: <http://pog.oxfordjournals.org/content/70/5/646.short>, acedido em 14 de Maio de 2012
- Särndal, C.-E., Swensson, B., Wretman, J. (1992). Model Assisted Survey Sampling. Springer-Verlag. New York
- Steeh, C. G. (1981). Trends in Nonresponse Rates, 1952-1979. *Public Opinion Quarterly*, Vol. 45, 40-57. Artigo disponível no sítio: <http://www.jstor.org/stable/2748317>

The American Association for Public Opinion Research. (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition. AAPOR. Artigo disponível no sítio: <http://www.aapor.org>, acedido em 26 de Julho de 2012

Wagner, J. (2010). The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey data. *Public Opinion Quarterly*, Vol. 74, Nº 2, 223–243. DOI: 10.1093/Poq/Nfq007

ANEXOS

A1 – Indicadores Básicos de Moçambique derivados dos Censos de 1997 e 2007

REGIÃO NORTE

Ord	INDICADORES BÁSICOS	NIASSA		CABO DELGADO		NAMPULA	
		ANO: 1997	2007	1997	2007	1997	2007
1	População urbana (em %)	23,1	22,9	16,8	20,8	25	28,6
2	População rural (em %)	76,9	77,1	83,2	79,2	75	71,4
3	Densidade demográfica (por Km2)	5,9	9,1	15,6	19,4	36,7	48,8
4	Taxa bruta de natalidade (por mil)	42,3	47,7	51,5	41,2	51,7	42,2
5	Taxa bruta de mortalidade (por mil)	21,9	14,5	23,8	16,5	23,2	13,4
6	Taxa de mortalidade infantil (por mil)	142,5	116,6	158,4	87,7	154	98,2
7	Taxa global de fecundidade (filhos/mulher)	6,8	6,9	6,7	5,7	7	5,8
8	Esperança de vida, total (em anos)	42,2	52,5	37,9	48,2	39,9	53,7
9	Esperança de vida, homens (em anos)	41,1	50,2	36,3	46,2	39,4	52,6
10	Esperança de vida, mulheres (em anos)	45,5	54,9	39,6	50,3	40,5	54,7
11	Taxa de Mortalidade Materna (por 100.000 nascimentos)	...	383,4	...	822,1	...	569,6
12	Taxa de analfabetismo, total (em %)		61	75	66,6	71,7	62,3
13	Taxa de analfabetismo, homens (em %)	52,2	44,6	60	50,8	56,7	46,5
14	Taxa de analfabetismo, mulheres (em %)	84,2	76,3	88,5	80,9	85,9	77,4
15	Taxa bruta de escolarização (em %) (b)						
	Ensino Primário do 1º Grau	48,3	88,8	43,5	95,9	47,5	91,9
	Ensino Primário do 2º Grau		72,3		76,3		61,1
	Ensino Secundário do 1º Ciclo	4,4	35,3	3,4	32,5	3,9	27,4
	Ensino Secundário do 2º Ciclo		14,1		12,1		12,5
	Ensino Superior	0	0,9	0	0,3	0	1,1
16	Número médio de pessoas por agregado familiar	4	4,3	3,8	4	3,7	4
17	Habitacões com electricidade (em %)	1,9	4,5	1,7	3,3	3,8	6
18	Habitacões segundo principal fonte de água para beber (em %):						
	Canalizada dentro de casa	0,4	0,5	0,6	0,6	0,9	0,7
	Canalizada fora de casa/quintal	2,5	2,3	2,9	4,4	5,7	6,2
	Fontanário	2,7	3,6	5	9,1	4,7	9,6
	Poço/furo protegido com bomba manual	63,1	16	76,2	15,7	75,1	7,9
	Poço sem bomba		51,2		53,6		55,7
	Rio/lago/lagoa	28,5	25,8	14,9	16,1	13,4	19,8
	Chuva		0,3		0,2		0
	Outra	2,8	0,3	0,4	0,3	0,1	0,2
19	Agregados Familiares segundo posse de bens duráveis (em %):						
	Rádio	22,0	51,2	20,8	42,8	20,3	46,4
	Televisor		3,7		3,0		4,0
	Telefone fixo		0,4		0,4		0,4
	Computador		0,2		0,2		0,3

A1 – Continuação

REGIÃO CENTRO

Ord	INDICADORES BÁSICOS	ZAMBÉZIA		TETE		MANICA		SOFALA	
		ANO: 1997	2007	1997	2007	1997	2007	1997	2007
1	População urbana (em %)	13,5	17,4	14,7	13,7	28,2	25,3	41,3	38,3
2	População rural (em %)	86,5	82,6	85,3	86,3	71,8	74,7	58,7	61,7
3	Densidade demográfica (por Km2)	27,5	36,7	11,4	17,7	15,8	22,9	20	24,2
4	Taxa bruta de natalidade (por mil)	50,0	46,1	47,9	46,3	48,8	46,5	43,5	41,7
5	Taxa bruta de mortalidade (por mil)	26,4	14,5	20,5	14	20,7	15,5	21	16,4
6	Taxa de mortalidade infantil (por mil)	168,6	101,6	129	82,6	139,1	100,1	139,6	96,5
7	Taxa global de fecundidade (filhos/mulher)	6,6	6,4	6,7	6,9	7	6,6	5,8	5,8
8	Esperança de vida, total (em anos)	37	52,3	43,8	51,4	42,7	49,1	42,2	47
9	Esperança de vida, homens (em anos)	36	49,9	42,3	49,4	40,7	46,8	40,4	45,1
10	Esperança de vida, mulheres (em anos)	38	54,8	45,2	53,5	44,9	51,5	44,1	49
11	Taxa de Mortalidade Materna (por 100.000 nascimentos)	...	518,6	...	271,7	...	389	...	656,4
12	Taxa de analfabetismo, total (em %)	70,3	62,5	66,8	56,2	57,7	43	56,2	43,4
13	Taxa de analfabetismo, homens (em %)	53,2	43,5	50	39,3	38,5	23,8	35,9	23
14	Taxa de analfabetismo, mulheres (em %)	85,2	79	81	71,5	73,9	59,6	74,8	61,9
15	Taxa bruta de escolarização (em %) (b)								
	Ensino Primário do 1º Grau	55,8	106,2	49,7	96,3	61,6	122,3	65,1	111
	Ensino Primário do 2º Grau		54,7		66,7		91,5		93
	Ensino Secundário do 1º Ciclo	3,2	26	5,2	28,7	5,2	43,7	9,5	46,5
	Ensino Secundário do 2º Ciclo		11,2		11,8		15,2		23,3
	Ensino Superior	0,0	0,6	0,0	0,3	0,1	0,7	0,2	3,3
16	Número médio de pessoas por agregado familiar	4,0	4,2	4,2	4,5	4,7	5,0	4,6	4,8
17	Habitações com electricidade (em %)	1,5	3,6	2,9	4,9	3,2	7,9	6,4	12,6
18	Habitações segundo principal fonte de água para beber (em %):								
	Canalizada dentro de casa	0,4	0,4	1,1	0,9	1,2	0,8	3,5	3,7
	Canalizada fora de casa/quintal	1,1	1,5	3,1	3,9	2,3	3,1	11,1	13,5
	Fontanário	3,6	5,9	7,4	6,5	6,3	5,6	7,6	13,3
	Poço/furo protegido com bomba manual	76,5	9,9	53,5	21,8	57,2	19,5	56,2	19,3
	Poço sem bomba		63,7		40		41		34,2
	Rio/lago/lagoa	17,1	17,1	34,2	26,5	32,4	29,1	20,6	15,5
	Chuva		0,0		0,1		0,1		0,2
	Outra	1,3	1,5	0,7	0,3	0,6	0,8	1,0	0,3
19	Agregados Familiares segundo posse de bens duráveis (em %):								
	Rádio	20,3	44,8	25,5	49,9	32,8	57,4	34,7	61,5
	Televisor		2,7		3,7		7,8		12,0
	Telefone fixo		0,3		0,4		0,6		1,2
	Computador		0,2		0,2		0,4		1,1

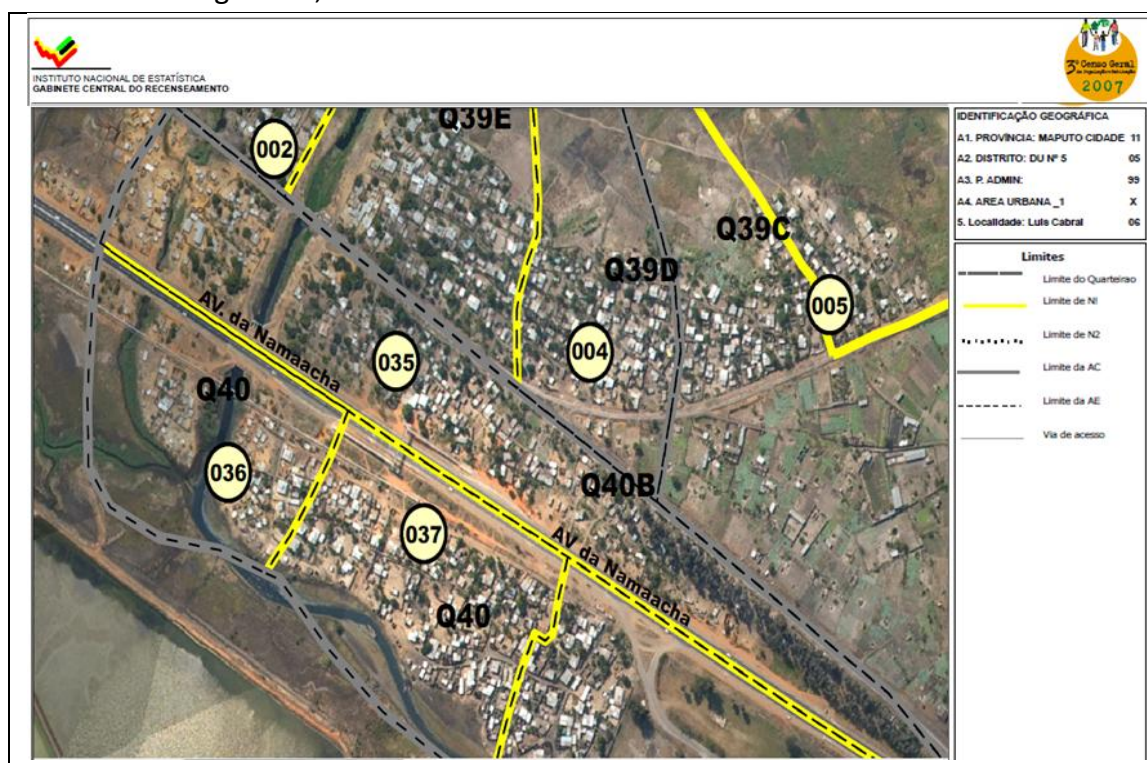
A1 – Continuação

REGIÃO SUL

Ord	INDICADORES BÁSICOS	INHAMBANE		GAZA		MAPUTO PROVÍNCIA		CIDADE DE MAPUTO	
		ANO: 1997	2007	1997	2007	1997	2007	1997	2007
1	População urbana (em %)	19,6	22,2	24,7	25,4		67,9	100	100
2	População rural (em %)	80,4	77,8	75,3	74,6		32,1	0,0	0,0
3	Densidade demográfica (por Km2)	16,4	18,5	14	16,2	30,9	46,3	3222,8	3648,8
4	Taxa bruta de natalidade (por mil)	42,9	39,5	44	40,3	37,7	33,2	35,5	27,1
5	Taxa bruta de mortalidade (por mil)	19,9	15,2	20,1	20,7	14,7	11,9	8,4	10,1
6	Taxa de mortalidade infantil (por mil)	118,4	80,4	124,4	89	97,2	70,6	79,7	61,6
7	Taxa global de fecundidade (filhos/mulher)	5,5	5,1	5,9	5,3	5	3,9	4,2	2,96
8	Esperança de vida, total (em anos)	46	51,8	45,6	43,5	50,6	54,6	58,3	56,2
9	Esperança de vida, homens (em anos)	42	48,7	41,1	40,7	47,2	51,9	55,1	53,2
10	Esperança de vida, mulheres (em anos)	49,3	54,9	49,2	46,2	53,9	57,3	61,8	59,1
11	Taxa de Mortalidade Materna (por 100.000 nascimentos)	...	567,7	...	388	...	330,5	...	364,1
12	Taxa de analfabetismo, total (em %)	54,2	41,3	52,7	38,5	34,3	22	15	9,8
13	Taxa de analfabetismo, homens (em %)	35,1	24,3	35,8	23,5	20,2	12,1	7,1	4,4
14	Taxa de analfabetismo, mulheres (em %)	66,4	52,7	63	48,8	45,9	30,5	22,6	14,8
15	Taxa bruta de escolarização (em %) (b)								
	Ensino Primário do 1º Grau	89,4	130,1	97,1	132,4	110,1	133,2	132,8	131,5
	Ensino Primário do 2º Grau		91,5		90,4		124,5		148,5
	Ensino Secundário do 1º Ciclo	4,4	46,8	4,5	44,3	11,5	74,9	26,6	114
	Ensino Secundário do 2º Ciclo		16,6		13,8		29,3		58,4
	Ensino Superior	0	0,9	0,1	1,3	0,8	5,7	3,4	14,2
16	Número médio de pessoas por agregado familiar	4,2	4,4	4,5	4,9	4,4	4,5	5,3	4,9
17	Habitacões com electricidade (em %)	1,7	4,9	4,8	12,3	12,1	29,2	38,1	63,0
18	Habitacões segundo principal fonte de água para beber (em %):								
	Canalizada dentro de casa	1,0	1,0	2,0	1,7	9	5,5	22,1	16,0
	Canalizada fora de casa/quintal	2,7	3,9	8,2	10,7	22,2	32	26,9	39,1
	Fontanário	3,7	6,7	10,1	20,4	18,2	19,1	26,6	33,3
	Poço/furo protegido com bomba manual	81	23,1	64,1	22,5	37,2	13,5	23,9	3,8
	Poço sem bomba		49,0		31,4		20,8		5,8
	Rio/lago/lagoa	7,2	9,0	13,6	10,4	12,1	7,2		0,0
	Chuva		6,5		2		0,2		0,0
	Outra	4,3	0,7	2	0,9	1,4	1,8	0,5	1,0
19	Agregados Familiares segundo posse de bens duráveis (em %):								
	Rádio	31,6	48,3	36,4	48,9	48,1	58,9	71,2	68,4
	Televisor		11,9		15,0		35,7		62,3
	Telefone fixo		0,7		0,6		1,5		6,2
	Computador		0,4		0,6		3,2		11,6

A2 – Modelo de mapa topográfico ou croquis de uma Área de Enumeração urbana

Neste mapa podem ser vistas, por exemplo, as áreas urbanas identificadas pelos códigos 035, 036 e 037 bem como seus limites.



A3 – Modelo de Ficha de listagem de AF's usada durante o IAF2002/03

IAF 2002/3

FICHA DE LISTAGEM DE AGREGADOS FAMILIARES

Página_____ de _____ páginas

PROVÍNCIA _____

DISTRITO _____

P. ADM _____

LOCALID _____

ALDEIA/BAIRRO _____

NOME DA ÁREA DE ENUMERAÇÃO _____

IDENTIFICAÇÃO ÁREA DE ENUMERAÇÃO (IAF I.D.) _____

AGENTE DE LISTAGEM _____

URBANO/RURAL _____

[illegible]

A4 – Taxas de desemprego em Moçambique segundo IAF2002/03 e IFTRAB2004

Estimativas da Taxa de Desemprego para pessoas de 15 anos e mais (INE, IAF2002/03)

Domínio	Estimativa (%)	Erro Padrão (%)	Coef. Variação (%)	Limites de Confiança, 95%		DEFF	Número de Observações
				Inferior (%)	Superior (%)		
Moçambique	3,1	0,2	6,90	2,7	3,5	3,10	24.262
Urbano	9,2	0,8	9,10	7,6	10,9	5,06	12.641
Rural	0,5	0,1	17,81	0,3	0,7	2,41	11.620
Niassa	0,1	0,0	63,89	0,0	0,2	0,30	2.060
Cabo Delgado	0,5	0,3	57,51	-0,1	1,2	3,51	1.624
Nampula	1,1	0,5	45,41	0,1	2,1	9,40	1.755
Zambézia	0,4	0,2	56,24	0,0	0,9	4,95	1.772
Tete	0,1	0,0	100,08	0,0	0,1	0,55	1.799
Manica	1,5	0,4	24,55	0,8	2,1	1,13	2.516
Sofala	4,8	1,1	22,44	2,7	6,9	4,14	2.339
Inhambane	0,9	0,3	32,92	0,3	1,6	1,61	1.957
Gaza	3,8	0,7	18,01	2,5	5,2	1,91	2.324
Maputo Província	13,4	1,1	8,49	11,2	15,6	1,27	2.535
Maputo Cidade	24,1	1,4	5,70	21,4	26,7	1,02	3.580

Estimativas da Taxa de Desemprego para pessoas de 15 anos e mais (INE, IFTRAB2004)

Domínio	Estimativa (%)	Erro Padrão (%)	Coef. Variação (%)	Limites de Confiança, 95%		DEFF	Número de Observações
				Inferior (%)	Superior (%)		
Moçambique	2,8	0,2	6,02	2,5	3,1	1,98	21.463
Urbano	7,9	0,4	5,12	7,1	8,7	1,39	12.157
Rural	0,3	0,1	22,87	0,2	0,5	2,20	9.306
Niassa	0,5	0,2	35,78	0,1	0,8	0,58	1.476
Cabo Delgado	0,9	0,3	31,65	0,4	1,5	1,56	1.998
Nampula	1,1	0,3	25,03	0,5	1,6	2,42	1.877
Zambézia	0,6	0,4	63,59	-0,2	1,4	9,34	1.801
Tete	0,3	0,1	28,79	0,1	0,4	0,33	1.961
Manica	1,0	0,3	30,42	0,4	1,6	1,13	1.625
Sofala	2,5	0,8	29,98	1,0	4,0	3,60	2.050
Inhambane	1,4	0,4	31,40	0,5	2,2	2,04	1.979
Gaza	3,6	0,7	19,40	2,2	5	1,74	1.878
Maputo Província	11,9	0,9	7,67	10,1	13,6	0,93	2.178
Maputo Cidade	17,1	0,9	5,50	15,3	18,9	0,70	2.640

A5 – Teste Binomial para Taxa de Resposta do IDS2003

Resultado

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Completo	730	77,6	77,6	77,6
	Ausente	171	18,2	18,2	95,7
	Recusa Total	18	1,9	1,9	97,7
	Incompleta	3	,3	,3	98,0
	Incapacitada	14	1,5	1,5	99,5
	Outro	5	,5	,5	100,0
	Total	941	100,0	100,0	

Número total de visitas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	707	75,1	75,1	75,1
	2	96	10,2	10,2	85,3
	3	138	14,7	14,7	100,0
	Total	941	100,0	100,0	

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Entrevista completa	Group 1	Sim	730	,8	,9	,000 ^{a, b}
	Group 2	Não	211	,2		
	Total		941	1,0		

a. Alternative hypothesis states that the proportion of cases in the first group < ,9.

b. Based on Z Approximation.

A6 – Teste de Qui-quadrado para Fatores da não-resposta do IDS2003

Resultado

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ausente	171	81,0	81,0	81,0
	Recusa Total	18	8,5	8,5	89,6
	Incompleta	3	1,4	1,4	91,0
	Incapacitada	14	6,6	6,6	97,6
	Outro	5	2,4	2,4	100,0
	Total	211	100,0	100,0	

Chi-Square Test

Resultado

	Observed N	Expected N	Residual
Ausente	171	42,2	128,8
Recusa Total	18	42,2	-24,2
Incompleta	3	42,2	-39,2
Incapacitada	14	42,2	-28,2
Outro	5	42,2	-37,2
Total	211		

Test Statistics

	Resultado
Chi-square	495,043 ^a
df	4
Asymp. Sig.	,000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 42,2.

A7 – Teste Binomial para Taxa de resposta do IFTRAB2004

Resultado					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Completo	1570	90,2	90,2	90,2
	Incompleto	14	,8	,8	91,0
	Todo o agregado familiar ausente	45	2,6	2,6	93,6
	Recusa durante a entrevista	2	,1	,1	93,7
	Recusa total	104	6,0	6,0	99,7
	Outro	5	,3	,3	100,0
	Total	1740	100,0	100,0	

Visitas				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1587	91,2	91,2
	2	83	4,8	96,0
	3	68	3,9	99,9
	4	2	,1	100,0
	Total	1740	100,0	

Binomial Test						
			Category	N	Observed Prop.	Test Prop.
Entrevista completa no AF	Group 1	Sim	1570	,9		
	Group 2	Não	170	,1		
	Total		1740	1,0		
						Asymp. Sig. (1-tailed)
						,394 ^a

a. Based on Z Approximation.

A8 – Teste de Qui-quadrado para Fatores de não-resposta do IFTRAB2004

Resultado					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Incompleto	14	8,2	8,2	8,2
	Todo o agregado familiar ausente	45	26,5	26,5	34,7
	Recusa durante a entrevista	2	1,2	1,2	35,9
	Recusa total	104	61,2	61,2	97,1
	Outro	5	2,9	2,9	100,0
	Total	170	100,0	100,0	

Chi-Square Test

Resultado			
	Observed N	Expected N	Residual
Incompleto	14	34,0	-20,0
Todo o agregado familiar ausente	45	34,0	11,0
Recusa durante a entrevista	2	34,0	-32,0
Recusa total	104	34,0	70,0
Outro	5	34,0	-29,0
Total	170		

Test Statistics	
	Resultado
Chi-square	214,294 ^a
df	4
Asymp. Sig.	,000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 34,0.

A9 – Teste Binomial para Taxa de resposta do IOF2009

visitas				
		Frequency	Percent	Cumulative Percent
Valid	1	1088	90,7	90,7
	2	86	7,2	97,9
	3	22	1,8	99,7
	4	2	,2	99,9
	5	1	,1	100,0
Total		1199	100,0	

resultado					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Completo	1052	87,7	87,7	87,7
	Incompleto	5	,4	,4	88,2
	Todo o AF ausente	24	2,0	2,0	90,2
	Recusa durante a entrevista	26	2,2	2,2	92,3
	Recusa total	74	6,2	6,2	98,5
	AF incapacitado	7	,6	,6	99,1
	Outro	11	,9	,9	100,0
Total		1199	100,0	100,0	

Binomial Test						
		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Entrevista completa no AF	Group 1	≤ 1	1052	,9	,9	,006 ^{a,b}
	Group 2	> 1	147	,1		
	Total		1199	1,0		

a. Alternative hypothesis states that the proportion of cases in the first group \leq ,9.

b. Based on Z Approximation.

A10 – Teste de Qui-quadrado para Fatores de não-resposta do IOF2009

resultado					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Incompleto	5	3,4	3,4	3,4
	Todo o AF ausente	24	16,3	16,3	19,7
	Recusa durante a entrevista	26	17,7	17,7	37,4
	Recusa total	74	50,3	50,3	87,8
	AF incapacitado	7	4,8	4,8	92,5
	Outro	11	7,5	7,5	100,0
Total		147	100,0	100,0	

Chi-Square Test

resultado			
	Observed N	Expected N	Residual
Incompleto	5	24,5	-19,5
Todo o AF ausente	24	24,5	-,5
Recusa durante a entrevista	26	24,5	1,5
Recusa total	74	24,5	49,5
AF incapacitado	7	24,5	-17,5
Outro	11	24,5	-13,5
Total	147		

Test Statistics	
	resultado
Chi-square	135,571 ^a
df	5
Asymp. Sig.	,000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 24,5.

A10.1 – Teste de Qui-quadrado para dois Fatores específicos da não-resposta do IOF2009

Chi-Square Test

Frequencies				
	resultado			
	Category	Observed N	Expected N	Residual
1	Todo o AF ausente	24	25,0	-1,0
2	Recusa durante a entrevista	26	25,0	1,0
Total		50		

Test Statistics	
	resultado
Chi-square	,080 ^a
df	1
Asymp. Sig.	,777

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25,0.

A11 – Teste de independência de Qui-quadrado entre Fatores de Não-resposta e Tipo de inquérito

Resultado * Tipo de inquérito Crosstabulation						
Resultado			Tipo de inquérito			Total
			IDS2003	IFTRAB2004	IOF2009	
Resultado	Ausente	Count	171	45	24	240
		Expected Count	95,9	77,3	66,8	240,0
		% within Tipo de inquérito	81,0%	26,5%	16,3%	45,5%
	Recusa Total	Count	18	104	74	196
		Expected Count	78,3	63,1	54,6	196,0
		% within Tipo de inquérito	8,5%	61,2%	50,3%	37,1%
	Recusa Durante a entrevista	Count	0	2	26	28
		Expected Count	11,2	9,0	7,8	28,0
		% within Tipo de inquérito	,0%	1,2%	17,7%	5,3%
Resultado	Incompleta	Count	3	14	5	22
		Expected Count	8,8	7,1	6,1	22,0
		% within Tipo de inquérito	1,4%	8,2%	3,4%	4,2%
	Incapacitada	Count	14	0	7	21
		Expected Count	8,4	6,8	5,8	21,0
		% within Tipo de inquérito	6,6%	,0%	4,8%	4,0%
	Outro	Count	5	5	11	21
		Expected Count	8,4	6,8	5,8	21,0
		% within Tipo de inquérito	2,4%	2,9%	7,5%	4,0%
Total	Count		211	170	147	528
	Expected Count		211,0	170,0	147,0	528,0
	% within Tipo de inquérito		100,0%	100,0%	100,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	266,636 ^a	10	,000
Likelihood Ratio	285,782	10	,000
Linear-by-Linear Association	53,833	1	,000
N of Valid Cases	528		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,85.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,711	,000
	Cramer's V	,502	,000
	Contingency Coefficient	,579	,000
N of Valid Cases		528	

A12 – Níveis de significância dos fatores identificados por inquérito

Fatores de Não-Resposta										
Tipo de inquérito			Estimate	Standard Error	95% Confidence Interval		Coefficient of Variation	Design Effect	Square Root Design Effect	Unweighted Count
					Lower	Upper				
IDS2003	% of Total	Ausente	81,0%	2,6%	75,4%	85,6%	,032	.	.	171
		Recusa Total	8,5%	1,9%	5,4%	13,2%	,228	.	.	18
		Incompleta	1,4%	,8%	,5%	4,3%	,564	.	.	3
		Incapacitada	6,6%	1,9%	3,7%	11,5%	,287	.	.	14
		Outro	2,4%	1,0%	1,0%	5,4%	,427	.	.	5
		Total	100,0%	,0%	100,0%	100,0%	,000	.	.	211
IFTRAB2004	% of Total	Ausente	26,5%	5,2%	17,5%	38,0%	,198	.	.	45
		Recusa Total	61,2%	6,5%	47,9%	73,0%	,106	.	.	104
		Recusa Durante a entrevista	1,2%	,8%	,3%	4,7%	,715	.	.	2
		Incompleta	8,2%	2,6%	4,4%	14,9%	,312	.	.	14
		Outro	2,9%	1,4%	1,2%	7,3%	,467	.	.	5
		Total	100,0%	,0%	100,0%	100,0%	,000	.	.	170
IOF2009	% of Total	Ausente	16,3%	4,7%	9,0%	27,8%	,287	.	.	24
		Recusa Total	50,3%	7,5%	35,9%	64,8%	,150	.	.	74
		Recusa Durante a entrevista	17,7%	5,8%	8,9%	32,0%	,328	.	.	26
		Incompleta	3,4%	1,7%	1,2%	9,0%	,503	.	.	5
		Incapacitada	4,8%	2,1%	1,9%	11,3%	,451	.	.	7
		Outro	7,5%	3,4%	2,9%	17,7%	,459	.	.	11
	Total	100,0%	,0%	100,0%	100,0%	,000	.	.	147	